(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 628 078 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
08.12.1999 Bulletin 1999/49

(21) Application number: **94903587.7**

(22) Date of filing: **10.12.1993**

(51) Int Cl.[6]: **C12N 15/13**, C07K 16/30, C07K 14/46, C12N 15/62, A61K 39/395

(86) International application number:
PCT/US93/12039

(87) International publication number:
WO 94/13806 (23.06.1994 Gazette 1994/14)

(54) **MULTIVALENT SINGLE CHAIN ANTIBODIES**

Multivalente einkettige Antikörper

ANTICORPS MONOCATENAIRES MULTIVALENTS

(84) Designated Contracting States:
AT BE CH DE FR GB IT LI LU NL SE

(30) Priority: **11.12.1992 US 990263**

(43) Date of publication of application:
14.12.1994 Bulletin 1994/50

(73) Proprietor: **THE DOW CHEMICAL COMPANY**
Midland, Michigan 48674 (US)

(72) Inventors:
- **MEZES, Peter, S.**
  Oldlyme, CT 06371 (US)
- **GOURLIE, Brian, B.**
  Midland, MI 48640 (US)

(74) Representative:
Weiss, Wolfgang, Dipl.-Chem. Dr. et al
Patentanwälte
Weickmann & Partner,
Kopernikusstrasse 9
81679 München (DE)

(56) References cited:
EP-A- 0 506 124
WO-A-91/19739
WO-A-93/11161

- **CANCER RESEARCH vol. 52, no. 12 , 15 June 1992 , PHILADELPHIA, PA, USA pages 3402 - 3408 T.YOKATA ET AL. 'Rapid tumour penetration of a single-chain Fv and comparison with other immunoglobulin forms'**
- **BIOCHEMISTRY vol. 30, no. 42 , 22 October 1991 , EASTON, PA US pages 10117 - 10125 M.W.PANTOLIANO ET AL. 'Conformational stability, folding and ligand-binding affinity of single-chain Fv immunoglobulin fragments expressed in Escherichia coli' cited in the application**

EP 0 628 078 B1

Description

[0001] The present invention relates to single chain multivalent antibodies.

[0002] Antibodies are proteins belonging to a group of immunoglobulins elicited by the immune system in response to a specific antigen or substance which the body deems foreign. There are five classes of human antibodies, each class having the same basic structure. The basic structure of an antibody is a tetramer, or a multiple thereof, composed of two identical heterodimers each consisting of a light and a heavy chain. The light chain is composed of one variable (V) and one constant (C) domain, while a heavy chain is composed of one variable and three or more constant domains. The variable domains from both the light and heavy chain, designated V_L and V_H respectively, determine the specificity of an immunoglobulin, while the constant (C) domains carry out various effector functions.

[0003] Amino acid sequence data indicate that each variable domain comprises three complementarity determining regions (CDR) flanked by four relatively conserved framework regions (FR). The FR are thought to maintain the structural integrity of the variable region domain. The CDR have been assumed to be responsible for the binding specificity of individual antibodies and to account for the diversity of binding of antibodies.

[0004] As the basic structure of an anti body contains two heterodimers, antibodies are multivalent molecules. For example, the IgG classes have two identical antigen binding sites, while the pentameric IgM class has 10 identical binding sites.

[0005] Monoclonal antibodies having identical genetic parentage and binding specificity have been useful both as diagnostic and therapeutic agents. Monoclonal antibodies are routinely produced by hybridomas generated by fusion of mouse lymphoid cells with an appropriate mouse myeloma cell line according to established procedures. The administration of murine antibodies for in vivo therapy and diagnostics in humans is limited however, due to the human anti-mouse antibody response illicited by the human immune system.

[0006] Chimeric antibodies, in which the binding or variable regions of antibodies derived from one species are combined with the constant regions of antibodies derived from a different species, have been produced by recombinant DNA methodology. See, for example, Sahagen et al.,*J. Immunol.,* 137 : 1066-1074(1986); Sun et al., *Proc. Natl. Acad. Sci. USA,* 82: 214-218 (1987) ; Nishimura et al., *Cancer Res.,* 47:999-1005 (1987) ; and Lie et al. *Proc Natl. Acad. Sci. USA,* 84: 3439-3443 (1987) which disclose chimeric antibodies to tumor-associated antigens. Typically, the variable region of a murine anti body is joined with the constant region of a human antibody. It is expected that as such chimeric antibodies are largely human in composition, they will be substantially less immunogenic than murine antibodies.

[0007] Chimeric antibodies still carry the Fc regions which are not necessary for antigen binding, but constitute a major portion of the overall antibody structure which affects its pharmacokinetics. For the use of antibodies in immunotherapy or immunodiagnostics, is it desirable to have antibody-like molecules which localize and bind to the target tissue rapidly and for the unbound material to quickly clear from the body. Generally, smaller antibody fragments have greater capillary permeability and are more rapidly cleared from the body than whole antibodies.

[0008] Since it is the variable regions of light and heavy chains that interact with an antigen, single chain antibody fragments (scFvs) have been created with one V_L and one V_H, containing all six CDR's, joined by a peptide linker (U. S. Patent 4,946,778) to create a V_L-L-V_H polypeptide, wherein the L stands for the peptide linker. A scFv wherein the V_L and V_H domains are orientated V_H-L-V_L is disclosed in U.S. Patent 5,132,405.

[0009] As the scFvs have one binding site as compared to the minimum of two for complete antibodies, the scFvs have reduced avidity as compared to the antibody containing two or more binding sites.

[0010] It would therefore be advantageous to obtain constructions of scFvs having more than one binding site to enhance the avidity of the polypeptide, and retain or increase their antigen recognition properties. In addition, it would be beneficial to obtain multivalent scFvs which are bispecific to allow for recognition of different epitopes on the target tissue, to allow for antibody-based recruitment of other immune effector functions, or allow antibody capture of a therapeutic or diagnostic moiety.

[0011] It has been found that single chain antibody fragments, each having one V_H and one V_L domain covalently linked by a first peptide linker, can be covalently linked by a second peptide linker to form a multivalent single chain antibody which maintains the binding affinity of a whole antibody. In one embodiment, the present invention is a multivalent single chain antibody having affinity for an antigen wherein the multivalent single chain antibody comprises two or more light chain variable domains and two or more heavy chain variable domains; wherein, each variable domain is linked to at least one other variable domain.

[0012] In another embodiment, the present invention is a multivalent single chain antibody which comprises two or more single chain antibody fragments, each single chain antibody fragment specifically binding an antigen, wherein the single chain antibody fragments are covalently linked by a first peptide linker which contains an amino acid sequence of

Leu Ser Ala Asp Asp Ala Lys Lys Asp Ala Ala Lys Lys Asp
Asp Ala Lys Lys Asp Asp Ala Lys Lys Asp Leu

and each single chain antibody fragment comprises

(a) a first polypeptide comprising a light chain variable domain;
(b) a second polypeptide comprising a heavy chain variable domain; and
(c) a second peptide linker linking the first and second polypeptides into a functional binding moiety.

[0013] In another embodiment, the invention provides a DNA sequence which codes for a multivalent single chain antibody, the multivalent single chain antibody comprising two or more single chain antibody fragments, each fragment having affinity for an antigen, wherein the fragments are covalently linked by a first peptide linker which contains an amino acid sequence of

Leu Ser Ala Asp Asp Ala Lys Lys Asp Ala Ala Lys Lys Asp Asp Ala Lys
Lys Asp Asp Ala Lys Lys Asp Leu

and each fragment comprising

(a) a first polypeptide comprising a light chain variable domain;
(b) a second polypeptide comprising a heavy chain variable domain; and
(c) a second peptide linker linking the first and second polypeptides into a functional binding moiety.

[0014] The multivalent single chain antibodies allow for the construction of an antibody fragment which has the specificity and avidity of a whole antibody but are smaller in size allowing for more rapid capillary permeability. Multivalent single chain antibodies also allow for the construction of a multivalent single chain antibody wherein the binding sites can be two different antigenic determinants.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] Figure 1 illustrates covalently linked single chain antibodies having the configuration V_L-L-V_H-L-V_L-L-V_H (LHLH) and V_L-L-V_H-L-V_H-L-V_L (LHHL) and a noncovalently linked Fv single chain antibody (Fv2).
[0016] Figure 2 illustrates the nucleotide sequence of CC49 V_L (SEQ ID NO: 1).
[0017] Figure 3 illustrates the amino acid sequence of CC49 V_L (SEQ ID NO: 2).
[0018] Figure 4 illustrates the nucleotide sequence of CC49 V_H (SEQ ID NO: 3) .
[0019] Figure 5 illustrates the amino acid sequence of CC49 V_H (SEQ ID NO:4).
[0020] Figure 6 illustrates the nucleotide sequence and amino acid sequence of the CC49 single chain antibody LHLH in p49LHLH (SEQ ID NO:6).
[0021] Figure 7 illustrates the nucleotide sequence and amino acid sequence of the CC49 single antibody LHHL in p49LHHL (SEQ ID NO:8).
[0022] Figure 8 illustrates construction of plasmids pSL301 T and pSL301 HT.
[0023] Figure 9 illustrates construction of plasmid p49LHHL.
[0024] Figure 10 illustrates construction of plasmid p49LHLH.
[0025] Figure 11 illustrates the results of a competition assay using CC49 IgG, CC49 scFv2, and CC49 scFv using biotinylated CC49 IgG as competitor.
[0026] The entire teaching of all references cited herein are hereby incorporated by reference.
[0027] Nucleic acids, amino acids, peptides, protective groups, active groups and such, when abbreviated, are abbreviated according to the IUPAC IUB (Commission on Biological Nomenclature) or the practice in the fields concerned.
[0028] The term "single chain antibody fragment" (scFv) or "antibody fragment" as used herein means a polypeptide containing a V_L domain linked to a V_H domain by a peptide linker (L), represented by V_L-L-V_H. The order of the V_L and V_H domains can be reversed to obtain polypeptides represented as V_H-L-V_L. "Domain" is a segment of protein that assumes a discrete function, such as antigen binding or antigen recognition.

[0029] A "multivalent single chain antibody" means two or more single chain antibody fragments covalently linked by a peptide linker. The antibody fragments can be joined to form bivalent single chain anti bodies having the order of the V_L and V_H domains as follows:

V_L-L-V_H-L-V_L-L-V_H; V_L-L-V_H-L-V_H-L-V_L; V_H-L-V_L-L-V_H-L-V_L; or V_H-L-V_L-L-V_L-L-V_H.

Single chain multivalent antibodies which are trivalent and greater have one or more anti body fragments joined to a bivalent single chain antibody by an additional interpeptide linker. In a preferred embodiment, the number of V_L and V_H domains is equivalent.

[0030] The present invention also provides for multivalent single chain antibodies which can be designated V_H-L-V_H-L-V_L-L-V_L or V_L-L-V_L-L-V_H-L-V_H.

[0031] Covalently linked single chain antibodies having the configuration V_L-L-V_H-L-V_L-L--V_H (LHLH) and V_L-L-V_H-L-V_H-L-V_L (LHHL) are illustrated in Figure 1. A noncovalently linked Fv single chain antibody (Fv2) is also illustrated in Figure 1.

[0032] The single chain antibody fragments for use in the present invention can be derived from the light and/or heavy chain variable domains of any antibody. Preferably, the light and heavy chain variable domains are specific for the same antigen. The individual antibody fragments which are joined to form a multivalent single chain antibody may be directed against the same antigen or can be directed against different antigens.

[0033] To prepare a vector containing the DNA sequence for a single chain multivalent antibody, a source of the genes encoding for these regions is required. The appropriate DNA sequence can be obtained from published sources or can be obtained by standard procedures known in the art. For example, Kabat et al., *Sequences of Proteins of Immunological Interest 4th ed.,* (1991), published by The U.S. Department of Health and Human Services, discloses sequences of most of the anti body variable regions which have been described to date.

[0034] When the genetic sequence is unknown, it is generally possible to utilize c DNA sequences obtained from mRNA by reverse transcriptase mediated synthesis as a source of DNA to clone into a vector. For antibodies, the source of mRNA can be obtained from a wide range of hybridomas. See, for example, the catalogue ATCC Cell Lines and Hybridomas, American Type Culture Collection, 20309 Parklawn Drive, Rockville Md., USA (1990). Hybridomas secreting monoclonal antibodies reactive with a wide variety of antigens are listed therein, are available from the collection, and usable in the present invention. These cell lines and others of similar nature can be utilized as a source of mRNA coding for the variable domains or to obtain antibody protein to determine amino acid sequence of the monoclonal antibody itself.

[0035] Variable regions of antibodies can also be derived by immunizing an appropriate vertebrate, normally a domestic animal, and most conveniently a mouse. The immunogen will be the antigen of interest, or where a hapten, an antigenic conjugate of the hapten to an antigen such as keyhole limpet hemocyanin (KLH). The immunization may be carried out conventionally with one or more repeated injections of the immunogen into the host mammal, normally at two to three week intervals. Usually, three days after the last challenge, the spleen is removed and dissociated into single cells to be used for cell fusion to provide hybridomas from which mRNA can readily be obtained by standard procedures known in the art.

[0036] When an antibody of interest is obtained, and only its amino acid sequence is known, it is possible to reverse translate the sequence.

[0037] The V_L and V_H domains for use in the present invention are preferably obtained from one of a series of CC antibodies against tumor-associated glycoprotein 72 antigen (TAG-72) disclosed in published PCT Applicadon WO 90/04410 on May 3, 1990, and published PCT Application WO 89/00692 on January 26, 1989. More preferred are the V_L and V_H domains from the monoclonal antibody designated CC49 in PCT Publications WO 90/04410 and WO 89/00692. The nucleotide sequence (SEQ ID NO: 1) which codes for the V_L of CC49 is substantially the same as that given in Figure 2. The amino acid sequence (SEQ ID NO: 2) of the V_L of CC49 is substantially the same as that given in Figure 3. The nucleotide sequence (SEQ ID NO: 3) which codes for the V_H of CC49 is substantially the same as that given in Figure 4. The amino acid sequence (SEQ ID NO: 4) for the V_H of CC49 is substantially the same as that given in Figure 5.

[0038] To form the antibody fragments and multivalent single chain antibodies of the present invention, it is necessary to have a suitable peptide linker. Suitable linkers for joining the V_H and V_L domains are those which allow the V_H and V_L domains to fold into a single polypeptide chain which will have a three dimensional structure very similar to the original structure of a whole antibody and thus maintain the binding specificity of the whole antibody from which antibody fragment is derived. Suitable linkers for linking the scFvs are those which allow the linking of two or more scFvs such that the V_H and V_L domains of each immunoglobulin fragment have a three dimensional structure such that each fragment maintains the binding specificity of the whole antibody from which the immunoglobulin fragment is derived. Linkers having the desired properties can be obtained by the method disclosed in U.S. Patent 4,946,778, the disclosure of which is hereby incorporated by reference. From the polypeptide sequences generated by the methods described in the 4,946,778, genetic sequences coding for the polypeptide can be obtained.

[0039] Preferably, the peptide linker joining the V_H and V_L domains to form a scFv and the peptide linker joining two

or more scFvs to form a multivalent single chain antibody have substantially the same amino acid sequence.

[0040] It is also necessary that the linker peptides be attached to the antibody fragments such that the binding of the linker to the individual antibody fragments does not interfere with the binding capacity of the antigen recognition site.

[0041] A preferred linker is based on the helical linker designated 205C as disclosed in Pantoliano et al. *Biochem.*, 30 10117-10125 (1991) but with the first and last amino acids changed because of the codon dictated by the Xho I site at one end and the Hind III site at the other. The amino acid sequence (SEQ ID NO: 5) of the preferred linker is as follows:

Leu-Ser-Ala-Asp-Asp-Ala-Lys-Lys-Asp-Ala-Ala-Lys-Lys-Asp-Asp-Ala-Lys-Lys-Asp-Asp-Ala-
-Lys-Lys-Asp-Leu.

[0042] The linker is generally 10 to 50 amino acid residues. Preferably, the linker is 10 to 30 amino acid residues. More preferably the linker is 12 to 30 amino acid residues. Most preferred is a linker of 15 to 25 amino acid residues.

[0043] Expression vehicles for production of the molecules of the invention include plasmids or other vectors. In general, such vectors contain replicon and control sequences which are derived from species compatible with a host cell. The vector ordinarily carries a replicon site, as well as specific genes which are capable of providing phenotypic selection in transformed cells. For example, *E. coli* is readily transformed using pBR322 [Bolivar et al., *Gene*, 2, 95-(1977), or Sambrook et al., *Molecular Cloning,* Cold Spring Harbor Press, New York, 2nd Ed. (1989)].

[0044] Plasmids suitable for eukaryotic cells may also be used. *S. cerevisiae,* or common baker's yeast, is the most commonly used among eukaryotic microorganisms, although a number of other strains, such as Pichia pastoris, are available. Cultures of cells derived from multicellular organisms such as SP2/0 or Chinese Hamster Ovary (CHO), which are available from the ATCC, may also be used as hosts. Typical of vector plasmids suitable for mammalian cells are pSV2neo and pSV2gpt (ATCC); pSVL and pKSV-10 (Pharmacia), pBPV-1/pML2d (International Biotechnology, Inc.).

[0045] The use of prokaryotic and eukaryotic viral expression vectors to express the genes for polypeptides of the present invention is also contemplated.

[0046] It is preferred that the expression vectors and the inserts which code for the single chain multivalent antibodies have compatible restriction sites at the insertion junctions and that those restriction sites are unique to the areas of insertion. Both vector and insert are treated with restriction endonucleases and then ligated by any of a variety of methods such as those described in Sambrook et al., supra.

[0047] Preferred genetic constructions of vectors for production of single chain multivalent antibodies of the present invention are those which contain a constitutively active transcriptional promoter, a region encoding signal peptide which will direct synthesis/secretion of the nascent single chain polypeptide out of the cell. Preferably, the expression rate is commensurate with the transport, folding and assembly steps to avoid accumulation of the polypeptide as insoluble material. In addition to the replicon and control sequences, additional elements may also be needed for optimal synthesis of single chain polypeptide. These elements may include splice signals, as well as transcription promoter, enhancers, and termination signals. Furthermore, additional genes and their products may be required to facilitate assembly and folding (chaperones).

[0048] Vectors which are commercially available can easily be altered to meet the above criteria for a vector. Such alterations are easily performed by those of ordinary skill in the art in light of the available literature and the teachings herein.

[0049] Additionally, it is preferred that the cloning vector contain a selectable marker, such as a drug resistance marker or other marker which causes expression of a selectable trait by the host cell. "Host cell" refers to cells which can be recombinantly transformed with vectors constructed using recombinant DNA techniques. A drug resistance or other selectable marker is intended in part to facilitate in the selection of transformants. Additionally, the presence of a selectable marker, such as a drug resistance marker, may be of use in keeping contaminating microorganisms from multiplying in the culture medium. In this embodiment, such a pure culture of the transformed host cell would be obtained by culturing the cells under conditions which require the induced phenotype for survival.

[0050] Recovery and purification of the present invention can be accomplished using standard techniques known in the art. For example, if they are secreted into the culture medium, the single chain multivalent antibodies can be concentrated by ultrafiltration. When the polypeptides are transported to the periplasmic space of a host cell, purification can be accomplished by osmotically shocking the cells, and proceeding with ultrafiltration, antigen affinity column chromatography or column chromatography using ion exchange chromatography and gel filtration. Polypeptides which are insoluble and present as refractile bodies, also called inclusion bodies, can be purified by lysis of the cells, repeated centrifugation and washing to isolate the inclusion bodies, solubilization, such as with guanidine-HCl, and refolding followed by purification of the biologically active molecules.

[0051] The activity of single chain multivalent antibodies can be measured by standard assays known in the art, for example competition assays, enzyme-linked immunosorbant assay (ELISA), and radioimmunoassay (RIA).

[0052] The multivalent single chain antibodies of the present invention provide unique benefits for use in diagnostics and therapeutics. The use of multivalent single chain antibodies afford a number of advantages over the use of larger fragments or entire antibody molecules. They reach their target tissue more rapidly, and are cleared more quickly from the body.

[0053] For diagnostic and/or therapeutic uses, the multivalent single chain antibodies can be constructed such that one or more anti body fragments are directed against a target tissue and one or more anti body fragments are directed against a diagnostic or therapeutic agent.

[0054] The invention also concerns pharmaceutical compositions which are particularly advantageous for use in the diagnosis and/or therapy of diseases, such as cancer, where target antigens are often expressed on the surface of cells. For diagnostic and/or therapeutic uses, the multivalent single chain antibodies can be conjugated with an appropriate imaging or therapeutic agent by methods known in the art. The pharmaceutical compositions of the invention are prepared by methods known in the art, e.g., by conventional mixing, dissolving or lyophilizing processes.

[0055] The invention will be further clarified by a consideration of the following examples, which are intended to be purely exemplary of the present invention.

ABBREVIATIONS

[0056]

BCIP	5-bromo-4-chloro-3-indoyl phosphate
bp	base pair
Bis-Tris propane	(1,3-bis[tris(hydroxymethyl)-methylamino]propane)
BSA	bovine serum albumin
CDR	Complementarity determining region
ELISA	enzyme linked immunosorbent assay
Fv2	non-covalent single chain Fv dimer
IEF	isoelectric focusing
Kbp	kilo base pair
LB	Luria-Bertani medium
Mab	monoclonal antibody
MES	2-(N-Morpholino)ethane sulfonic acid
MW	molecular weight
NBT	nitro blue tetrazolium chloride
Oligo	Oligonucleotides
PAG	polyacrylamide gel
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pSCFV	plasmid containing DNA sequence coding for SCFV
RIGS	radioimmunoguided surgery
RIT	radioimmunotherapy
scFv	single chain Fv immunoglobulin fragment monomer
scFv2	single chain Fv immunoglobulin fragment dimer covalently linked
SDS	sodium dodecyl sulfate
TBS	Tris-buffered saline
Tris	(Tris[hydroxymethyl]aminomethane)
TTBS	Tween-20 wash solution
V_H	immunoglobulin heavy chain variable domain
V_L	immunoglobulin light chain variable domain

Antibodies

[0057] CC49: A murine monoclonal antibody specific to the human tumor-associated glycoprotein 72 (TAG-72) deposited as ATCC No. HB9459.

[0058] CC49 FAB: An antigen binding portion of CC49 consisting of an intact light chain linked to the N-terminal portion of the heavy chain.

[0059] CC49 scFv: Single chain antibody fragment consisting of two variable domains of CC49 antibody joined by a peptide linker.

[0060] CC49 Fv2: Two CC49 scFv non-covalently linked to form a dimer. The number after Fv refers to the number of monomer subunits of a given molecule, e.g., CC49 Fv6 refers to the hexamer multimers.
[0061] CC49 scFv2: Covalently-linked single chain antibody fragment consisting of two CC49 V_L domains and two VH domains joined by three linkers. Six possible combinations for the order of linking the V_L(L) and the V_H(H) domains together are: LHLH, LHHL, LLHH, HLLH, HLHL, and HHLL.

Plasmids

[0062] pSCFV UHM: Plasmid containing coding sequence for scFv consisting of a CC49 variable light chain and a CC49 variable heavy chain joined by a 25 amino acid linker.
[0063] p49LHLH or p49LHHL: Plasmids containing the coding sequence for producing CC49 scFv2 LHLH or LHHL products, respectively.

EXAMPLES

General Experimental

[0064] Procedures for molecular cloning are as those described in Sambrook et al., *Molecular Cloning,* Cold Spring Harbor Press, New York, 2nd Ed. (1989) and Ausubel et al., *Current Protocols in Molecular Biology,* John Wiley and Sons, New York (1992), the disclosures of which are hereby incorporated by reference.
[0065] All water used throughout was deionized distilled water.

Oligonucleotide Synthesis and Purification

[0066] All oligonuclotides (oligos) were synthesized on either a Model 380A or a Model 391 DNA Synthesizer from Applied Biosystems (Foster City, CA) using standard β-cyanoethyl phosphoramidites and synthesis columns. Protecting groups on the product were removed by heating in concentrated ammonium hydroxide at 55°C for 6 to 15 hours. The ammonium hydroxide was removed through evaporation and the crude mixtures were resuspended in 30 to 40 μL of sterile water. After electrophoresis on polyacrylamide-urea gels, the oligos were visualized using short wavelength ultraviolet (UV) light. DNA bands were excised from the gel and eluted into 1 mL of 100 mM Tris-HCl, pH 7.4, 500 mM NaCl, 5 mM EDTA over 2 hours at 65°C. Final purification was achieved by applying the DNA to Sep-Pac™ C-18 columns (Millipore, Bedford, MA) and eluting the bound oligos with 60 percent methanol. The solution volume was reduced to approximately 50 μL and the DNA concentration was determined by measuring the optical density at 260 nm (OD_{260}).

Restriction Enzyme Digests

[0067] All restriction enzyme digests were performed using Bethesda Research Laboratories (Gaithersburg, MD), New England Biolabs, Inc. (Beverly, MA) or Boehringer Mannheim (BM, Indianapolis, IN) enzymes and buffers following the manufacturer's recommended procedures. Digested products were separated by polyacrylamide gel electrophoresis (PAGE). The gels were stained with ethidium bromide, the DNA bands were visualized using long wavelength UV light and the DNA bands were then excised. The gel slices were placed In dialysis tubing (Union Carbide Corp., Chicago) containing 5 mM Tris, 2.5 mM acetic acid, 1 mM EDTA, pH 8.0 and eluted using a Max Submarine electrophoresis apparatus (Hoefer Scientific Instruments, CA). Sample volumes were reduced on a Speed Vac Concentrator (Savant Instruments, Inc., NY). The DNA was ethanol precipitated and redissolved in sterile water.

Enzyme Linked Immunosorbent Assay (ELISA)

[0068] TAG-72 antigen, prepared substantially as described by Johnson et al, *Can. Res.,* 46, 850-857 (1986), was adsorbed onto the wells of a polyvinyl chloride 96 well microtiter plate (Dynatech Laboratories, Inc., Chantilly, VA) by drying overnight. The plate was blocked with 1 percent BSA in PBS for 1 hour at 31°C and then washed 3 times with 200 μL of PBS, 0.05 percent Tween-20. 25 μL of test antibodies and 25 μL of biotinylated CC49 (1/20,000 dilution of a 1 mg/mL solution) were added to the wells and the plate incubated for 30 minutes at 31°C. The relative amounts of TAG-72 bound to the plate, biotinylated CC49, streptavidinalkaline phosphatase, and color development times were determined empirically in order not to have excess of either antigen or biotinylated CC49, yet have enough signal to detect competition by scFv. Positive controls were CC49 at 5 μg/mL and CC49 Fab at 10 μg/mL. Negative controls were 1 percent BSA in PBS and/or concentrated LB. Unbound proteins were washed away. 50 μL of a 1:1000 dilution of streptavidin conjugated with alkaline phosphatase (Southern Biotechnology Associates, Inc., Birmingham, AL) were

added and the plate was incubated for 30 minutes at 31°C. The plate was washed 3 more times. 50 µL of a para-nitrophenyl-phosphate solution (Kirkegaard & Perry Laboratories, Inc., Gaithersburg, MD) were added and the color reaction was allowed to develop for a minimum of 20 minutes. The relative amount of scFv2 binding was measured by optical density scanning at 404-450 nm using a microplate reader (Molecular Devices Corporation, Manlo Park, CA). Binding of the scFv2 species resulted in decreased binding of the biotinylated CC49 with a concomitant decrease in color development.

SDS-PAGE and Western Blotting

[0069] Samples for SDS-PAGE analysis (20 µL) were prepared by boiling in a non-reducing sample preparation buffer-Seprasol I (Integrated Separation Systems (ISS), Natick, MA) for 5 minutes and loaded on 10-20 percent gradient polyacrylamide Daiichi Minigels as per the manufacturer's directions (ISS).

[0070] Electrophoresis was conducted using a Mini 2-gel apparatus (ISS) at 55 mA per gel at constant current for approximately 75 minutes. Gels were stained in Coomassie Brilliant Blue R-250 (Bio-Rad, Richmond, CA) for at least 1 hour and destained. Molecular weight standards were prestained (Mid Range Kit, Diversified Biotech, Newton Center, MA) and included the following proteins: Phosphorylase b, glutamate dehydrogenase, ovalbumin, lactate dehydrogenase, carbonic amhydrase, B-lactoglobulin and cytochrome C. The corresponding MWs are: 95,500, 55,000, 43,000, 36,000, 29,000, 18,400, and 12,400, respectively.

[0071] When Western analyses were conducted, a duplicate gel was also run. After electrophoresis, one of the gels was equilibrated for 15-20 minutes in anode buffer #1 (0.3 M Tris-HCl pH 10.4). An Immobilon-P PVDF (polyvinylidene dichlorine) membrane (Millipore, Bedford, MA) was treated with methanol for 2 seconds, and immersed in water for 2 minutes. The membrane was then equilibrated in anode buffer #1 for 3 minutes. A Milliblot-SDE apparatus (Millipore) was utilized to transfer proteins in the gel to the membrane. A drop of anode buffer #1 was placed in the middle of the anode electrode surface. A sheet of Whatman 3MM filter paper was soaked in anode buffer #1 and smoothly placed on the electrode surface. Another filter paper soaked in anode buffer #2 (25 mM tris pH 10.4) was placed on top of the first one. A sandwich was made by next adding the wetted PVDF membrane, placing the equilibrated gel on top of this and finally adding a sheet of filter paper soaked in cathode buffer (25mM Tris-HCl, pH 9.4 in 40 mM glycine). Transfer was accomplished in 30 minutes using 250 mA constant current (initial voltage ranged from 8-20 volts).

[0072] After blotting, the membrane was rinsed briefly in water and placed in a dish with 20 mL blocking solution (1 percent bovine serum albumin (BSA) (Sigma, St. Louis, MO) in Tris-buffered saline (TBS)). TBS was purchased from Pierce Chemical (Rockford, IL) as a preweighed powder such that when 500 mL water is added, the mixture gives a 25 mM Tris, 0.15 M sodium chloride solution at pH 7.6. The membranes were blocked for a minimum of 1 hour at ambient temperature and then washed 3 times for 5 minutes each using 20 mL 0.5 percent Tween-20 wash solution (TTBS). To prepare the TTBS, 0.5mL of Tween 20 (Sigma) was mixed per liter of TBS. The probe antibody used was 20 m L bioti nylated FAID14 solution (10 µg per 20 mL antibody buffer). Antibody buffer was made by adding 1 g BSA per 100 mL of TTBS. After probing for 30-60 minutes at ambient temperature, the membrane was washed 3 times with TTBS, as above.

[0073] Next, the membrane was incubated for 30-60 minutes at ambient temperature with 20 mL of a 1:500 dilution in antibody buffer of streptavidin conjugated with alkaline phosphatase (Southern Biotechnology Associates, Birmingham, AL). The wash step was again repeated after this, as above. Prior to the color reaction, membranes were washed for 2 minutes in an alkaline carbonate buffer (20 mL). This buffer is 0.1 M sodium bicarbonate, 1 mM $MgCl_2 \cdot H_2O$, pH 9.8. To make up the substrate for alkaline phosphatase, nitroblue tetrazolium (NBT) chloride (50 mg, Sigma) was dissolved in 70 percent dimethylformamide. 5-Bromo-4-chloro-3-indoyl phosphate (BCIP) (25 mg, Sigma) was separately dissolved in 100 percent dimethylformamide. 5-Bromo-4-chloro-3-indoyl phosphate (BCIP) 25 mg, Sigma) was separately dissolved in 100 percent dimethylformamide. These solutions are also commercially available as a Western developing agent sold by Promega. For color development, 120 µL of each were added to the alkaline solution above and allowed to react for 15 minutes before they were washed from the developed membranes with water.

Biotinylated FAID14

[0074] FAID14 is a murine anti-idiotypic antibody (IgG2a, K isotype) deposited as ATCC No. CRL 10256 directed against CC49. FAID14 was purified using a Nygene Protein A affinity column (Yonkers, NY). The manufacturer's protocol was followed, except that 0.1 M sodium citrate, pH 3.0 was used as the elution buffer. Fractions were neutralized to pH-7 using 1.0 M Tris-HCl pH 9.0. The biotinylation reaction was set up as follows. FAID14 (1 mg, 100 µL in water) was mixed with 100 µL of 0.1 M Na_2CO_3 pH 9.6. Biotinyl-ε-amino-caproic acid N-hydroxy succinimide ester (Biotin-X-NHS) (Cal biochem, LaJolla, CA) (2.5 mg) was dissolved in 0.5 mL dimethylsulfoxide. Biotin-X-NHS solution (20 µL) was added to the FAID14 solution and allowed to react at 22°C for 4 hours. Excess biotin and impurities were removed by gel filtration, using a Pharmacia Superose 12 HR10/30 column (Piscataway, NJ). At a flow rate of 0.8 mL/min, the

biotinylated FAID14 emerged with a peak at 16.8 min. The fractions making up this peak were pooled and stored at 4°C and used to detect the CC49 idiotype as determined by the CC49 V_L and V_H CDRs.

Isoelectric Focusing (IEF)

[0075] Isoelectric points (pI's) were predicted using a computer program called PROTEIN-TITRATE, available through DNASTAR (Madison, WI). Based on amino acid composition with an input sequence, a MW value is given, in addition to the pI. Since Cys residues contribute to the charge, the count was adjusted to 0 for Cys, since they are all involved in disulfide bonds.

[0076] Experimentally, pI's were determined using Isogel agarose IEF plates, pH range 3-10 (FMC Bioproducts, Rockland, ME). A Biorad Bio-phoresis horizontal electrophoresis cell was used to run the IEF, following the directions of both manufacturers. The electrophoresis conditions were : 500 volts (limiting), at 20 mA current and 10 W of constant power. Focusing was complete in 90 min. IEF standards were purchased from Biorad; the kit included phycocyanin, β-lactoglobulin B, bovine carbonic anhydrase, human carbonic anhydrase, equine myoglobin, human hemoglobins A and C, 3 lentil lectins and cytochrome C, with pI values of 4.65, 5.10, 6.00, 6.50, 7.00, 7.10 and 7.50, 7.80, 8.00, and 8.20 and 9.60, respectively. Gels were stained and destained according to the directions provided by FMC.

Quantitation of CC49 Antibody Species

[0077] All purified CC49 antibodies including the IgG, scFv2 species and the monomeric scFv were quantitated by measuring the absorbence of protein dilutions at 280 mm using matching 1.0 cm pathlength quartz cuvettes (Hellma) and a Perkin-Elmer UV/VIS Spectrophotometer, Model 552A. Molar absorptivities (E_m) were determined for each antibody by using the following formula:

$$E_m = (\text{number Trp}) \times 5{,}500 + (\text{number Tyr}) \times 1{,}340 + (\text{number (Cys)2}) \times 150 + (\text{number Phe}) \times 10$$

The values are based on information given by D. B. Wetlaufer, *Advances in Protein Chemistry,* 17, 375-378).

High Performance Liquid Chromatography

[0078] All high performance liquid chromatography (HPLC) was performed for CC49 scFv2 purification using an LKB HPLC system with titanium or teflon tubing throughout. The system consists of the Model 2150 HPLC pump, model 2152 controller, UV CORD S11 model 2238 detection system set at an absorbence of 276 nm and the model 2211 SuperRac fraction collector.

PCR Generation of Subunits

[0079] All polymerase chain reactions (PCR) were performed with a reaction mixture consisting of: 150 picograms (pg) plasmid target (pSCFVUHM); 100 pmoles primers; 1 µL Perkin-Elmer-Cetus (PEC, Norwalk, CT) Ampli-Taq polymerase; 16 µL of 10 mM dNTPs and 10 µL of 10X buffer both supplied in the PEC kit; and sufficient water to bring the volume to total volume to 100 µL. The PCR reactions were carried out essentially as described by the manufacturer. Reactions were done in a PEC 9600 thermocycler with 30 cycles of: denaturation of the DNA at 94°C for 20 to 45 sec, annealing from between 52 to 60°C for 0.5 to 1.5 min., and elongation at 72°C for 0.5 to 2.0 min. Oligonucleotide primers were synthesized on an Applied Biosystems (Foster City, CA) 380A or 391 DNA synthesizer and purified as above.

Ligations

[0080] Ligation reactions using 100 ng of vector DNA and a corresponding 1 : 1 stoichiometric equivalent of insert DNA were performed using a Stratagene (La Jolla, CA) T4 DNA ligase kit following the manufacturer's directions. Ligation reactions (20 µL total volume) were initially incubated at 18°C and allowed to cool gradually overnight to 4°C.

Transformations

[0081] Transformations were performed utilizing 100 µL of Stratagene E. coli AG 1 competent cell (Stratagene, La

Jolla, CA) according to the directions provided by the manufacturer. DNA from the ligation reactions (1-5 μL) were used. After the transformation step, cells were allowed to recover for 1 hr in Luria broth (LB) at 37°C with continuous mixing and subsequently plated onto either 20 μg/mL chloramphenicol containing (CAM 20) Luria agar for pSCFVUHM, p49LHLH or p49LHHL or 100 μg/mL ampicillin (AMP 100) Luria agar plates (LB-AMP 100) for clones contai ning the plasmid pSL301 or subsequent constructions derived from pSL301.

Screening of E. coli Clones

[0082] Bacterial plasmids were isolated from LB broth culture containing the appropriate drug to maintain selection pressure using Promega (Madison, WI) Magic mini-prep plasmid preparation kits. The kit was used per the manufacturer's specifications.

Plasmid Constructions

[0083] Two plasmids, designated p49LHLH and p49LHHL, were constructed to produce multivalent single chain antibodies. The host cell containing p49LHLH produced a polypeptide which can be designated by V_L-L-V_H-L-V_L-L-V_H where V_L and V_H are the light and heavy cahin variable regions of CC49 antibody and linker (L) is a 25 amino acid linker having the sequence (SEQ ID NO: 5).

Leu-Ser-Ala-Asp-Asp-Ala-Lys-Lys-Asp-Ala-Ala-Lys-Lys-Asp-Asp-Ala-Lys-Lys-Asp-

-Asp-Ala-Lys-Lys-Asp-Leu.

[0084] The host cell containing p49LHHL produced a polypeptide which can be designated by V_L-L-V_H-L-V_H-L-V_L where V_L and V_H are the light and heavy chain variable domains of the CC49 antibody and L is a peptide linker having the amino acid sequence indicated above.

[0085] The nucleotide sequence (SEQ ID NO: 6) and amino acid sequence (SEQ ID NO: 7) of the CC49 V_L-L-V_H-L-V_L-L-V_H (p49LHLH) are given in Figure 6. The nucleotide sequence (SEQ ID NO: 8) and amino acid sequence (SEQ ID NO: 9) of the CC49 V_L-L-V_H-L-V_H-L-V_L (p49LHHL) are given in Figure 7.

Construction of pSL301 HT

[0086] The construction of pSL301 HT is illustrated in Figure 8. The Bacillus lichiformis penicillinase P (penP) terminator sequence was removed from the plasmid designated pSCFV UHM by a 45 minute digest with Nhe I and BamH I, excised from a 4.5 percent polyacrylamide gel after electrophoresis, electroeluted, ethanol precipitated and ligated into the same sites in the similarly prepared vector: pSL301 (Invitrogen, San Diego, CA). A procedure for preparing pSCFV UHM is given is U.S. patent application Ser. No. 07/935,695 filed August 21, 1992, the disclosure of which is hereby incorporated by reference. In general, pSCFV UHM contains a nucleotide sequence for a penP promoter; a unique Nco 11 restriction site; CC49 V_L region; Hind III restriction site; a 25 amino acid linker; a unique a Xho I restriction site; CC49 V_H region; Nhe I restriction site; penP terminator; and BamH I restriction site (see, Figure 8). The penP promoter and terminator are described in Mezes, et al. (1983), *J. Biol. Chem.*, 258 11211-11218 (1983).

[0087] An aliquot of the ligation reaction (3 μL) was used to transform competent *E. coli* AG cells which were plated on LB-AMP100 agar plates and grown overnight. Potential clones containing the penP terminator insert were screened using a Pharmacia (Gaithersburg, MD) T7 Quickprime ^{32}P DNA labeling kit in conjunction with the microwave colony lysis procedure outlined in Buluwela et al., *Nucleic Acid Research,* 17, 452 (1989). The probe, which was the penP-Nhe I-BamH I terminator fragment itself was prepared and used according to the directions supplied with the Quickprime kit. A clone which was probe positive and which contained the 207 base pair inserts from a BamH I and Nhe I digest (base pairs (bp) 1958 to 2165, Figure 6) was designated pSL301 T and chosen to construct pSL301 HT which would contain the nudeotide sequence for CC49 V_H. The reason the Nhe I-BamH I penP terminator was placed into pSL301 was to eliminate the Eco47 III restriction endonuclease site present in the polylinker region between its Nhe I and BamH I sites. This was designed to accommodate the subsequent build-up of the V_L and V_H domains where the Eco47 III site needed to be unique for the placement of each successive V domain into the construction. As each V domain was added at the Eco47 III-Nhe I sites, the Eco47 III was destroyed in each case to make the next Eco47 III site coming in on the insert unique.

[0088] The V_H sequence was made by PCR with oligos 5' SCP1 and 3'oligo SCPS using pSCFV U HM as the target for PCR amplification. The DNA sequence for SCP1 (SEQ ID NO: 10) and SCPS (SEQ ID NO: 11) are as follows:

SCP1: 5'-TAAA CTC GAG GTT CAG TTG CAG CAG -3'

SCP5: 5'-TAAA GCT AGC ACCA AGC GCT TAG TGA GGA GAC GGT GAC TGA GGT-3'

The underlined portion indicates the endonuclease restriction sites.

[0089] The amplified V_H DNA was purified from a 4 percent PAG, electroeluted ethanol precipitated and dissolved in 20 μL water. The V_H sequence was digested with Xho I and Nhe I restriction enzymes and used as the insert with the pSL301 T vector which had been digested with the same restriction enzymes and subsequently purified. A standard ligation reaction was done and an aliquot (4 μL) used to transform competent E. coli AG1 cells. The transformed cells were plated onto LB AMP100 agar plates. Candidate dones were picked from a Nhe I and Xho I digest screen that revealed that the CC49V_H insert had been obtained.

[0090] DNA sequencing was performed to verify the sequence of the CC49V_H with United States Biochemical (USB) (Cleveland, Ohio) Sequence kit and sequencing primers pSL301SEQB (a 21 bp sequencing primer which annealed in the pSL301 vector 57 bp upstream from the Xho I site) and CC49VHP, revealed clones with the correct CC49V_H sequence in pSL301HT. This plasmid was used as the starting point in the construction of both pSL301-HHLT and pSL301-HLHT. The sequencing oligos used are shown here.

[0091] The nucleotide sequence of pSL301SEQ B (SEQ ID NO: 12) and CC49V_H (SEQ ID No: 13) are as follows:

pSL301SEQB: 5'-TCG TCC GAT TAG GCA AGC TTA-3'

CC49VHP: 5'-GAT GAT TTT AAA TAC AAT GAG-3'

Example 1 p49LHHL Construction

[0092] Using pSL301 HT (5 μg) as the starting material, it was digested with Eco47 III and Nhe I and the larger vector fragment was purified. A CC49V_H insert fragment was generated by PCR using SCP6B is the 5' oligo and SCPS as the 3' oligo. The nucleotide sequence (SEQ ID NO: 14) of SCP6B is as follows:

SCP6B: 5'- TAAA TGC GCA GAT GAC GCA AAG AAA GAC GCA GCT AAA AAA GAC GAT GCC AAA AAG GAT GAC GCC AAG AAA GAT CTT GAG GTT CAG TTG CAG CAG TCT-G'

The oligo SCP6B also contains part of the coding region for the linker (bp 8-76 of SEQ ID NO: 14). The portion of the oligo designed to anneal with the CC49VH target in pSCFV UHM is from bp77-90 in SEQ ID NO: 14.

[0093] The underlined sequence corresponds to the Fsp I site. The resulting PCR insert was purified, digested with Fsp I and Nhe I and used in a ligation reaction with the pSL301 HT Eco47 III-Nhe I vector (Figure 7). Competent E. coli AG1 cells were used for the transformation of this ligation reaction (3 μL) and were plated on LB-AMP100 agar plates. Two clones having the correct size Xho I-Nhe I insert representative of the pSL301 HHT product were sequenced with the oligo SQP1 and a single clone with the correct sequence (nucleotides 1124-1543 of Figure 7) was chosen for further construction. The nucleotide sequence of SQP1 (SEQ ID NO: 15 is as follows:

SQP1: 5'-TG ACT TTA TGT AAG ATG ATG T-3'

[0094] The final linker-V_L subunit (bp 1544-1963, Figure 7) was generated using the 5'oligo, SCP7b and the 3' oligo, SCP8a, using pSCFV UHM as the target for the PCR. The nucleotide sequence of SCP7b (SEQ ID NO: 16) is as follows:

SCP7b: 5'-TAAA TGC GCA GAT GAC GCA AAG AAA GAC GCA GCT AAA AAA GAC GAT GCC AAA AAG GAT GAC GCC AAG AAA GAT CTT GAC ATT GTG ATG TCA CAG TCT CC

The underlined nucleotides correspond to an Fsp I site. The nucleotide sequence of SCP8a (SEQ ID NO: 17) is as

follows:

SCP8a: 5'-TAAA GCT AGC TTT TTA CTT AAG CAC CAG CTT GGT CCC-3'

[0095] The first set of underlined nucleotides correspond to an Nhe I site, while the other corresponds to an Afl II site. Nucleotides B-76 of SCP70 code for the linker (nucleotides 1544-1612 of Figure 7) while nucleotides 77-99 which anneal to the V_L correspond to 1613-1635 of Figure 7. The primer SCP8a has a short tail at its 5' end, a Nhe I restriction site, a stop codon, an Afl II restriction site and the last 21 bases of the V_L. After Fsp I and Nhe I digestion, this resulting 420 bp insert was purified and ligated into the Nhe I and Eco47 III sites of the purified pSL301 HHT vector, candidate clones were screened with Nhe I and Xho I, the correct size insert verified and sequenced with 49LFR2(-) and SQP1 to confirm the newly inserted sequence in pSL301HHLT. The nucleotide sequence (SEQ ID NO: 18) is as follows:

49LFR2(-): 5'-CTG CTG GTA CCA GGC CAA G-3'

[0096] The plasmid pSL301HHLT was digested with Xho I and Nhe I, purified, and the resulting 1179 bp V_H-linker-V_H-linker-V_L segment ligated into pSCFV UHM, which had been cut with the same restriction enzymes and the larger vector fragment purified, to form p49LHHL. The ligation reaction (4 µL aliquot) was used to transform competent E. coli AG 1 cells (Stratagene) and plated onto LBCAM20 agar plates. A single clone which had a plasmid with the correct restriction enzyme map was selected to contain p49LHHL. The p49LHHL contains a penP promoter and a nucleotide sequence for the CC49 multivalent single chain anti body scFv2:

V_L-L-V_H-L-V_H-L-V_L or CC49 scFv2 (LHHL).

Example 2: p49LHLH Construction

[0097] The construction of p49LHLH is schematically represented in Figure 10. A linker-V_L subunit was generated with the 5' oligo SCP7b and the 3'oligo SCP9 (SEQ ID NO: 19).

SCP9: 5'-TAA AGC TAG CAC CAA GCG CTT AGT TTC AGC ACC AGC TTG GTC CCA G-3'

[0098] The SCP7b oligo(nucleotides 8-76) codes for the linker in Figure 6 (corresponding to nucleotides 1124-1192) and annealed to the pSCFV UHM target for the PCR (nucleotides 77-99) corresponding to nucleotides 1193-1215 of the V_L in Figure 6.

[0099] SCP9 has a Nhe I site (first underlined nucleotides) and an Eco47 III site (second underlined nucleotides) which are restriction sites needed for making the pSL301HLT ready to accept the next V domain. Nucleotides 18-23 of SCP9 correspond to nucleotides 1532-1537 of Figure 6 (coding for the first 2 amino acids of the linker). while nucleotides 24-46 correspond to nucleotides 1508-1531 of Figure 6 which was also the annealing region for SCP9 in the PCR. The plasmid pSL301 HT was digested with Eco47 III and Nhe I and the larger vector fragment was purified for ligation with the linker-CC49V_L DNA insert fragment from the PCR which had been treated with Fsp I and Nhe I and purified. The ligation mixture (3 µL) was used to transform E. coli AG 1 competent cells and one colony having the correct Xho I-Nhe I size fragment was sequenced using the oligo PENPTSEQ2. The nudeotide sequence (SEQ. ID NO.20) is as follows:

5'-TTG ATC ACC AAG TGA CTT TAT G-3'

[0100] The sequencing results indicated that there had been a PCR error and deletion in the resulting pSL301HT clone. A five base deletion, corresponding to nucleotides 1533-1537 as seen in Figure 6 had been obtained and nucleotide 1531 which should have been a T was actually a G, as determined from the DNA sequence data. The resulting sequence was

5'...G AAGC GCT T...etc.

where the underlined sequence fortuitously formed an Eco47 III site. The AGCGCT sequence in Figure 6, would correspond to nucleotides 1530, 1531, 1532, 1538, 1539 and 1540. This error was corrected in the next step, generating

pSL301 HLHT, by incorporating the 5 base deletion at the end of oligo SCP6C (SEQ ID NO: 21).

SCP6C: 5'-TAAGCGCTGATGATGCTAAGAAGGACGCCGCAAAAAA
GGACGACGCAAAAAAAGATGATGCAAAAAAGGATCTGG
AGGTTCAGTTGCAGCAGTCTGAC-3'

[0101] The underlined sequence in SCP6c corresponds to an Eco47 III site. SCP6C was used as the 5' oligo, with SCP10 as the 3' oligo in a PCR to generate a linker CC49 V_L segment. The nudeotide sequence (SEQ ID NO: 22) is as follows:

SCP10: 5'TTG TGC TAG CTT TTT ATG AGG AGA CGG TGA CTG AGG TT-3'

[0102] The underlined sequence in SCP10 corresponds to the Nhe I site found at nucleotides 1958-1963 in Figure 6. The PCR insert was digested this time only with Nhe I and purified. The vector (pSL301 HLT) was digested at the Eco47 III site (that had been formed) and Nhe I and purified. The insert and vector were ligated and an aliquot (3 µL) used to transform competent E. coli AG1 cells. This was plated on LB-AMP100 plates and candidate clones screened with Xho I and Nhe I. Three clones having the correct size DNA were obtained. Two of these clones were sequenced using the oligo 49VLCDR3(+) and SQP1. The nucleotide sequence (SEQ ID NO:23) of 49VLCDR3(+) is as follows:

49VLCDR3(+):
5'-CAG CAG TAT TAT AGC TAT-3'

[0103] One clone, with the correct sequence was obtained and the sequence from nucleotides 1533 to 1963 in Figure 6 were verified, giving a correct pSL301 HLHL clone.

[0104] To generate the final plasmid, p49LHLH for expression in E. coli, pSL301 HLHT (5 µg) was digested with Nhe I and Xho I, and the smaller insert fragment containing the V_H-L-V_L-L-V_H sequence purified. It was ligated with the larger purified vector fragment from a digest of pSCFV UHM (5 µg) with Xho I and Nhe I. An aliquot of the ligation mix (4 µL) was used to transform competent E. coli AG1 cells. The transformation mix was plated on LB-CAM20 plates, and a representative clone for p49 LHLH was selected on the basis of a correct restriction enzyme map (see Figure 10) and biological activity toward TAG-72.

Example 3: Purification of CC49 scFv2 LHLH and LHHL Covalently Linked Dimers

[0105] For the purification of the CC49 covalently linked single chain dimers. (scFv2), *E. coli* periplasmic fractions were prepared from 1.0 L overnight cultures of both p49LHLH and p49LHHL. Briefly, the culture was divided into 4 X 250 mL portions and centrifuged at 5,000 rpm for 10 minutes in a Sorvall GS-3 rotor. The pelleted cells were washed and resuspended in 100 mL each of 10 mM Tris-HCl pH 7.3 containing 30 mM NaCl. The cells were again pelleted and washed with a total of 100 mL 30 mM Tris-HCl pH 7.3 and pooled into one tube. To this, 100 mL of 30 mM Tris-HCl pH 7.3 containing 40 percent w/v sucrose and 2.0 mL of 10 mM EDTA pH 7.5 was added. The mixture was kept at room temperature, with occasional shaking, for 10 minutes. The hypertonic cells were then pelleted as before. In the next step, the shock, the pellet was quickly suspended in 20 mL ice cold 0.5 mM $MgCl_2$ and kept on ice for 10 minutes, with occasional shaking. The cells were pelleted as before and the supernatant containing the *E. coli* periplasmic fraction was clarified further by filtration through a 0.2 µm Nalge (Rochester, NY) filter apparatus and concentrated in Amicon (Danvers, MA) Centriprep 30 and Centricon 30 devices to a volume of less than 1.0 mL.

[0106] The concentrated peri plasmic shockates from either the p49LHLH or p49LHHL clones were injected onto a Pharmacia (Piscataway, NJ) Superdex 75 HR 10/30 HPLC column that had been equilibrated with PBS. At a flow rate of 0.5 mL/minute, the product of interest, as determined by competition ELISA, had emerged between 21 through 24 minutes. The active fractions were pooled, concentrated as before and dialyzed overnight using a system 500 Micro-dialyzer Unit (Pierce Chemical) against 20 mM Tris-HCl pH 7.6 with 3-4 changes of buffer and using an 8,000 MW cutoff membrane. The sample was injected on a Pharmacia Mono Q HR 5/5 anion exchange HPLC column. A gradient program using 20 mM Tris-HCl pH 7.6 as buffer A and the same solution plus 0.5 M NaCl as buffer B was employed at a flow rate of 1.5 mL/min. The products of interest in each case, as determined by competition ELISA, emerged from the column between 3 and 4 minutes. Analysis of the fractions at this point on duplicate SDS-PAGE gels, one stained

with Coomassie Brilliant Blue R-250 and the other transferred for Western analysis (using biotinylated FAID 14 as the probe antibody) revealed a single band at the calculated molecular weight for the scFv2 (LHLH or LHHL) species at 58,239 daltons. The active fractions were in each case concentrated, dialysed against 50 mM MES pH 5.8 overnight and injected on a Pharmacia Mono S HR 5/5 cation exchange column. The two fractions of interest from this purification step, as determined by SDS-PAGE and ELISA, fractions 5 and 6, eluted just before the start of the gradient, so they had not actually bound to the column. Fractions 5 and 6 were consequently pooled for future purification.

[0107] A Mono Q column was again run on the active Mono S fractions but the buffer used was 20 mM Tris-HCl, pH 8.0 and the flow rate was decreased to 0.8 mL/minute. The products emerged without binding, but the impurity left over from the Mono S was slightly more held up, so that separation did occur between 5 and 6 minutes. After this run, the products were homogeneous and were saved for further characterization.

Isoelectric Focusing

[0108] The isoelectric points (pl) of the constructs was predicted using the DNASTAR (Madison, WI) computer program Protein-titrate. Based on amino acid composition, a MW and pl value was calculated.

[0109] Experimentally, pls were determined using FMC Bioproducts (Rockland, ME) Isogel IEF plates, pH range 3-10. A Biorad (Richmond, CA) electrophoresis unit was used to run the IEF, following the directions of both manufacturers. The electrophoresis conditions were as follows: 500 V (limiting) at 20 mA and at 10 W of constant power. Focusing was complete in 90 minutes. Biorad IEF standards included phycocyanin, beta lactoglobulin B, bovine carbonic anhydrase, human carbonic anhydrase, equine myoglobulin, human hemoglobins A and C, 3 lentil lectin, and cytochrome C with pl value of 4.65, 5.10, 6.00, 6,50, 7.00, 7.50, 7.8, 8.00, 8.20 and 9.6, respectively. Gels were stained and destained according to directions provided by FMC. The DNASTAR program predicted values of 8.1 for the pl for both scFv2 species. A single, homogeneous band for the pure products was observed on the gel at pl values for both at 6.9.

[0110] Purified CC49 antibodies such as the IgG, scFv2 (LHLH and LHHL) were quantitated by measuring the absorbence spectrophotometrically at 280 nm. Molar absorbtivity values, ε_M, were determined for each using the formula cited above by Wetlaufer.

[0111] Based on the amino acid composition, the $E^{0.1\%}$ (280 nanometers) values for CC49 IgG, CC49 scFv2 LHLH, CC49 scFv2 LHHL and CC49 scfv were 1.49, 1.65, 1.65 and 1.71, respectively.

Example 4

[0112] Relative activities of the CC49 scFv2 species LHLH and LHHL, were compared with the IgG and a monomer scfv form with a FLAG peptide at the COOH terminus.

[0113] Percent competition was determined from the ELISA data by the following equation:

$$\frac{\text{Zero competition - sample reading (OD405-450 nm)}}{\text{zero competition - 100 percent competition}} \times 100$$

[0114] The "zero competition" value was determined by mixing (1 : 1) one percent BSA with the biotinylated CC49 (3 X 10-14 moles) while the 100 percent competition value was based on a 5 μg/mL sample of CC49 IgG mixed with the biotinylated CC49 IgG. The data are presented in Figure 11. Absorbence values for the samples were measured at 405 nm - 450 nm. The average of triplicate readings was used. Initially samples (25 pL) were applied to the TAG-72 coated microliter plates at 1.0 X 10-10 moles of binding sites/mL. Biotinylated CC49 (4 μg/μL diluted 1:20,000 - used 25 μL) diluted the samples by a factor of 2. Serial dilutions (1:2) were performed. Both forms of the scFv2 are approximately equivalent to the IgG (see Figure 11). In a separate experiment, a CC49 scFv monomer was compared to a Fab fragment, both of which are monovalent and these were also shown to be equivalent in their binding affinity for TAG-72. These results indicate that both forms of the covalently linked dimers have 2 fully functional antigen binding sites. This is the same increase in avidity as observed with the whole IgG, relative to a monomeric species.

[0115] These data also indicate that the scFv2 molecules, like their CC49 IgG parent are candidates for immunotherapeutic applications, but with the benefit of increased capillary permeability and more rapid biodistribution pharmacokinetics. The advantage should allow multiple injections of compounds of the present invention and give higher tumor:tissue ratios in immunotherapeutic treatment regimens for cancer treatment, relative to the existing IgG molecules.

[0116] Other embodiments of the invention will be apparent to those skilled in the art from a consideration of this specification or practice of the invention disclosed herein. It is intended that the specification and examples be considered as exemplary only, with the true scope and spirit of the invention being indicated by the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

[0117]

(i) APPLICANT: The Dow Chemical Company

(ii) TITLE OF INVENTION: MULTIVALENT SINGLE CHAIN ANTIBODIES

(iii) NUMBER OF SEQUENCES: 23

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Duane C. Ulmer
(B) STREET: P.O. Box 1967
(C) CITY: Midland
(D) STATE: MI
(E) COUNTRY: US
(F) ZIP: 48641-1967

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Ulmer, Duane C
(B) REGISTRATION NUMBER: 34,941
(C) REFERENCE/DOCKET NUMBER: 41,014-F

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (517) 636-8104

(2) INFORMATION FOR SEQ ID NO:1:

[0118]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 339 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID XO:1:

```
GACATTGTGA TGTCACAGTC TCCATCCTCC CTACCTGTGT CAGTTGGCGA GAAGGTTACT    60
TTGAGCTGCA AGTCCAGTCA GAGCCTTTTA TATAGTGGTA ATCAAAAGAA CTACTTGGCC   120
TGGTACCAGC AGAAACCAGG GCAGTCTCCT AAACTGCTGA TTTACTGGGC ATCCGCTAGG   180
GAATCTGGGG TCCCTGATCG CTTCACAGGC AGTGGATCTG GGACAGATTT CACTCTCTCC   240
ATCAGCAGTG TGAAGACTGA AGACCTGGCA GTTTATTACT GTCAGCAGTA TTATAGCTAT   300
CCCCTCACGT TCGGTGCTGG GACCAAGCTG GTGCTGAAG                          339
```

(2) INFORMATION FOR SEQ ID NO:2:

[0119]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 113 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Asp Ile Val Met Ser Gln Ser Pro Ser Ser Leu Pro Val Ser Val Gly
1               5                   10                  15

Glu Lys Val Thr Leu Ser Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser
            20                  25                  30

Gly Asn Gln Lys Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln
        35                  40                  45

Ser Pro Lys Leu Leu Ile Tyr Trp Ala Ser Ala Arg Glu Ser Gly Val
    50                  55                  60

Pro Asp Arg Phe Thr Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser
65                  70                  75                  80

Ile Ser Ser Val Lys Thr Glu Asp Leu Ala Val Tyr Tyr Cys Gln Gln
                85                  90                  95

Tyr Tyr Ser Tyr Pro Leu Thr Phe Gly Ala Gly Thr Lys Leu Val Leu
            100                 105                 110

Lys
```

(2) INFORMATION FOR SEQ ID NO:3:

[0120]

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 345 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
GAGGTTCAGT TGCAGCAGTC TGACGCTGAG TTGGTGAAAC CTGGGGCTTC AGTGAAGATT    60
TCCTGCAAGG CTTCTGGCTA CACCTTCACT GACCATGCAA TTCACTGGGT GAAACAGAAC   120
CCTGAACAGG GCCTGGAATG GATTGGATAT TTTTCTCCCG GAAATGATGA TTTTAAATAC   180
AATGAGAGGT TCAAGGGCAA GGCCACACTG ACTGCAGACA AATCCTCCAG CACTGCCTAC   240
GTGCAGCTCA ACAGCCTGAC ATCTGAGGAT TCTGCAGTGT ATTTCTGTAC AAGATCCCTG   300
AATATGGCCT ACTGGGGTCA AGGAACCTCA GTCACCGTCT CCTCA                   345
```

(2) INFORMATION FOR SEQ ID NO:4:

[0121]

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 115 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Glu Val Gln Leu Gln Gln Ser Asp Ala Glu Leu Val Lys Pro Gly Ala
1               5                   10                  15

Ser Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Asp His
            20                  25                  30

Ala Ile His Trp Val Lys Gln Asn Pro Glu Gln Gly Leu Glu Trp Ile
        35                  40                  45

Gly Tyr Phe Ser Pro Gly Asn Asp Asp Phe Lys Tyr Asn Glu Arg Phe
    50                  55                  60

Lys Gly Lys Ala Thr Leu Thr Ala Asp Lys Ser Ser Ser Thr Ala Tyr
65                  70                  75                  80
```

Val Gln Leu Asn Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys
85 90 95

Thr Arg Ser Leu Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser Val Thr
100 105 110

Val Ser Ser
115

(2) INFORMATION FOR SEQ ID NO:5:

[0122]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Leu Ser Ala Asp Asp Ala Lys Lys Asp Ala Ala Lys Lys Asp Asp Ala
1 5 10 15

Lys Lys Asp Asp Ala Lys Lys Asp Leu
20 25

(2) INFORMATION FOR SEQ ID NO:6:

[0123]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2165 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CTCATGTTTG ACAGCTTATC ATCGATGAAT TCCATCACTT CCCTCCGTTC ATTTGTCCCC 60

GGTGGAAACG AGGTCATCAT TTCCTTCCGA AAAAACGGTT GCATTTAAAT CTTACATATA 120

TAATACTTTC AAAGACTACA TTTGTAAGAT TTGATGTTTG AGTCGGCTGA AAGATCGTAC 180

GTACCAATTA TTGTTTCGTG ATTGTTCAAG CCATAACACT GTAGGGATAG TGGAAAGAGT 240

GCTTCATCTG GTTACGATCA ATCAAATATT CAAACGGAGG GAGACGATTT TGATGAAATA 300

CCTATTGCCT ACGGCAGCCG CTGGATTGTT ATTACTCGCT GCCCAACCAG CCATGGCCGA 360

CATTGTGATG TCACAGTCTC CATCCTCCCT ACCTGTGTCA GTTGGCGAGA AGGTTACTTT 420

GAGCTGCAAG TCCAGTCAGA GCCTTTTATA TAGTGGTAAT CAAAAGAACT ACTTGGCCTG 480

GTACCAGCAG AAACCAGGGC AGTCTCCTAA ACTGCTGATT TACTGGGCAT CCGCTAGGGA 540

ATCTGGGGTC CCTGATCGCT TCACAGGCAG TGGATCTGGG ACAGATTTCA CTCTCTCCAT 600

CAGCAGTGTG AAGACTGAAG ACCTGGCAGT TTATTACTGT CAGCAGTATT ATAGCTATCC 660

CCTCACGTTC GGTGCTGGGA CCAAGCTGGT GCTGAAGCTT AGTGCGGACG ATGCGAAAAA 720

GGATGCTGCG AAGAAGGATG ACGCTAAGAA AGACGATGCT AAAAAGGACC TCGAGGTTCA 780

GTTGCAGCAG TCTGACGCTG AGTTGGTGAA ACCTGGGGCT TCAGTGAAGA TTTCCTGCAA 840

GGCTTCTGGC TACACCTTCA CTGACCATGC AATTCACTGG GTGAAACAGA ACCCTGAACA 900

GGGCCTGGAA TGGATTGGAT ATTTTTCTCC CGGAAATGAT GATTTTAAAT ACAATGAGAG 960

GTTCAAGGGC AAGGCCACAC TGACTGCAGA CAAATCCTCC AGCACTGCCT ACGTGCAGCT 1020

CAACAGCCTG ACATCTGAGG ATTCTGCAGT GTATTTCTGT ACAAGATCCC TGAATATGGC 1080

CTACTGGGGT CAAGGAACCT CAGTCACCGT CTCCTCACTA AGCGCAGATG ACGCAAAGAA 1140

AGACGCAGCT AAAAAAGACG ATGCCAAAAA GGATGACGCC AAGAAAGATC TTGACATTGT 1200

GATGTCACAG TCTCCATCCT CCCTACCTGT GTCAGTTGGC GAGAAGGTTA CTTTGAGCTG 1260

CAAGTCCAGT CAGAGCCTTT TATATAGTGG TAATCAAAAG AACTACTTGG CCTGGTACCA 1320

GCAGAAACCA GGGCAGTCTC CTAAACTGCT GATTTACTGG GCATCCGCTA GGGAATCTGG 1380

GGTCCCTGAT CGCTTCACAG GCAGTGGATC TGGGACAGAT TTCACTCTCT CCATCAGCAG 1440

TGTGAAGACT GAAGACCTGG CAGTTTATTA CTGTCAGCAG TATTATAGCT ATCCCCTCAC 1500

GTTCGGTGCT GGGACCAAGC TGGTGCTGAA GCTAAGCGCT GATGATGCTA AGAAGGACGC 1560

CGCAAAAAAG GACGACGCAA AAAAGGATGA TGCAAAAAAG GATCTGGAGG TTCAGTTGCA 1620

GCAGTCTGAC GCTGAGTTGG TGAAACCTGG GGCTTCAGTG AAGATTTCCT GCAAGGCTTC 1680

TGGCTACACC TTCACTGACC ATGCAATTCA CTGGGTGAAA CAGAACCCTG AACAGGGCCT 1740

GGAATGGATT GGATATTTTT CTCCCGGAAA TGATGATTTT AAATACAATG AGAGGTTCAA 1800

GGGCAAGGCC ACACTGACTG CAGACAAATC CTCCAGCACT GCCTACGTGC AGCTCAACAG 1860

CCTGACATCT GAGGATTCTG CAGTGTATTT CTGTACAAGA TCCCTGAATA TGGCCTACTG 1920

GGGTCAAGGA ACCTCAGTCA CCGTCTCCTC ATAAAAAGCT AGCGATGAAT CCGTCAAAAC 1980

ATCATCTTAC ATAAAGTCAC TTGGTGATCA AGCTCATATC ATTGTCCGGC AATGGTGTGG 2040

GCTTTTTTTG TTTTCTATCT TTAAAGATCA TGTGAAGAAA AACGGGAAAA TCGGTCTGCG 2100

GGAAAGGACC GGGTTTTTGT CGAAATCATA GGCGAATGGG TTGGATTGTG ACAAAATTCG 2160

GATCC 2165

(2) INFORMATION FOR SEQ ID NO:7:

[0124]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 553 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: Protein
(B) LOCATION: 23

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala
        -20                 -15                 -10

Ala Gln Pro Ala Met Ala Asp Ile Val Met Ser Gln Ser Pro Ser Ser
    -5                  1               5                   10

Leu Pro Val Ser Val Gly Glu Lys Val Thr Leu Ser Cys Lys Ser Ser
                15                  20                  25

Gln Ser Leu Leu Tyr Ser Gly Asn Gln Lys Asn Tyr Leu Ala Trp Tyr
            30                  35                  40

Gln Gln Lys Pro Gly Gln Ser Pro Lys Leu Leu Ile Tyr Trp Ala Ser
        45                  50                  55

Ala Arg Glu Ser Gly Val Pro Asp Arg Phe Thr Gly Ser Gly Ser Gly
    60                  65                  70

Thr Asp Phe Thr Leu Ser Ile Ser Ser Val Lys Thr Glu Asp Leu Ala
75                  80                  85                  90

Val Tyr Tyr Cys Gln Gln Tyr Tyr Ser Tyr Pro Leu Thr Phe Gly Ala
                95                  100                 105
```

Gly Thr Lys Leu Val Leu Lys Leu Ser Ala Asp Asp Ala Lys Lys Asp
110 115 120

Ala Ala Lys Lys Asp Asp Ala Lys Lys Asp Asp Ala Lys Lys Asp Leu
125 130 135

Glu Val Gln Leu Gln Gln Ser Asp Ala Glu Leu Val Lys Pro Gly Ala
140 145 150

Ser Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Asp His
155 160 165 170

Ala Ile His Trp Val Lys Gln Asn Pro Glu Gln Gly Leu Glu Trp Ile
175 180 185

Gly Tyr Phe Ser Pro Gly Asn Asp Asp Phe Lys Tyr Asn Glu Arg Phe
190 195 200

Lys Gly Lys Ala Thr Leu Thr Ala Asp Lys Ser Ser Ser Thr Ala Tyr
205 210 215

Val Gln Leu Asn Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys
220 225 230

Thr Arg Ser Leu Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser Val Thr
235 240 245 250

Val Ser Ser Leu Ser Ala Asp Asp Ala Lys Lys Asp Ala Ala Lys Lys
255 260 265

Asp Asp Ala Lys Lys Asp Asp Ala Lys Lys Asp Leu Asp Ile Val Met
270 275 280

Ser Gln Ser Pro Ser Ser Leu Pro Val Ser Val Gly Glu Lys Val Thr
285 290 295

Leu Ser Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser Gly Asn Gln Lys
300 305 310

Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Lys Leu
315 320 325 330

Leu Ile Tyr Trp Ala Ser Ala Arg Glu Ser Gly Val Pro Asp Arg Phe
335 340 345

Thr Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser Ile Ser Ser Val
350 355 360

Lys Thr Glu Asp Leu Ala Val Tyr Tyr Cys Gln Gln Tyr Tyr Ser Tyr
365 370 375

Pro Leu Thr Phe Gly Ala Gly Thr Lys Leu Val Leu Lys Leu Ser Ala
380 385 390

Asp Asp Ala Lys Lys Asp Ala Ala Lys Lys Asp Asp Ala Lys Lys Asp
395 400 405 410

Asp Ala Lys Lys Asp Leu Glu Val Gln Leu Gln Gln Ser Asp Ala Glu
415 420 425

Leu Val Lys Pro Gly Ala Ser Val Lys Ile Ser Cys Lys Ala Ser Gly
430 435 440

Tyr Thr Phe Thr Asp His Ala Ile His Trp Val Lys Gln Asn Pro Glu
445 450 455

Gln Gly Leu Glu Trp Ile Gly Tyr Phe Ser Pro Gly Asn Asp Asp Phe
460 465 470

Lys Tyr Asn Glu Arg Phe Lys Gly Lys Ala Thr Leu Thr Ala Asp Lys
475 480 485 490

Ser Ser Ser Thr Ala Tyr Val Gln Leu Asn Ser Leu Thr Ser Glu Asp
495 500 505

Ser Ala Val Tyr Phe Cys Thr Arg Ser Leu Asn Met Ala Tyr Trp Gly
510 515 520

Gln Gly Thr Ser Val Thr Val Ser Ser
525 530

(2) INFORMATION FOR SEQ ID NO:8:

[0125]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2165 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CTCATGTTTG ACAGCTTATC ATCGATGAAT TCCATCACTT CCCTCCGTTC ATTTGTCCCC 60

GGTGGAAACG AGGTCATCAT TTCCTTCCGA AAAAACGGTT GCATTTAAAT CTTACATATA 120

TAATACTTTC AAAGACTACA TTTGTAAGAT TTGATGTTTG AGTCGGCTGA AAGATCGTAC 180

GTACCAATTA TTGTTTCGTG ATTGTTCAAG CCATAACACT GTAGGGATAG TGGAAAGAGT 240

GCTTCATCTG GTTACGATCA ATCAAATATT CAAACGGAGG GAGACGATTT TGATGAAATA 300

```
CCTATTGCCT ACGGCAGCCG CTGGATTGTT ATTACTCGCT GCCCAACCAG CCATGGCCGA   360
CATTGTGATG TCACAGTCTC CATCCTCCCT ACCTGTGTCA GTTGGCGAGA AGGTTACTTT   420
GAGCTGCAAG TCCAGTCAGA GCCTTTTATA TAGTGGTAAT CAAAAGAACT ACTTGGCCTG   480
GTACCAGCAG AAACCAGGGC AGTCTCCTAA ACTGCTGATT TACTGGGCAT CCGCTAGGGA   540
ATCTGGGGTC CCTGATCGCT TCACAGGCAG TGGATCTGGG ACAGATTTCA CTCTCTCCAT   600
CAGCAGTGTG AAGACTGAAG ACCTGGCAGT TTATTACTGT CAGCAGTATT ATAGCTATCC   660
CCTCACGTTC GGTGCTGGGA CCAAGCTGGT GCTGAAGCTT AGTGCGGACG ATGCGAAAAA   720
GGATGCTGCG AAGAAGGATG ACGCTAAGAA AGACGATGCT AAAAAGGACC TCGAGGTTCA   780
GTTGCAGCAG TCTGACGCTG AGTTGGTGAA ACCTGGGGCT TCAGTGAAGA TTTCCTGCAA   840
GGCTTCTGGC TACACCTTCA CTGACCATGC AATTCACTGG GTGAAACAGA ACCCTGAACA   900
GGGCCTGGAA TGGATTGGAT ATTTTTCTCC CGGAAATGAT GATTTTAAAT ACAATGAGAG   960
GTTCAAGGGC AAGGCCACAC TGACTGCAGA CAAATCCTCC AGCACTGCCT ACGTGCAGCT  1020
CAACAGCCTG ACATCTGAGG ATTCTGCAGT GTATTTCTGT ACAAGATCCC TGAATATGGC  1080
CTACTGGGGT CAAGGAACCT CAGTCACCGT CTCCTCACTA AGCGCAGATG ACGCAAAGAA  1140
AGACGCAGCT AAAAAAGACG ATGCCAAAAA GGATGACGCC AAGAAAGATC TTGAGGTTCA  1200
GTTGCAGCAG TCTGACGCTG AGTTGGTGAA ACCTGGGGCT TCAGTGAAGA TTTCCTGCAA  1260
GGCTTCTGGC TACACCTTCA CTGACCATGC AATTCACTGG GTGAAACAGA ACCCTGAACA  1320
GGGCCTGGAA TGGATTGGAT ATTTTTCTCC CGGAAATGAT GATTTTAAAT ACAATGAGAG  1380
GTTCAAGGGC AAGGCCACAC TGACTGCAGA CAAATCCTCC AGCACTGCCT ACGTGCAGCT  1440
CAACAGCCTG ACATCTGAGG ATTCTGCAGT GTATTTCTGT ACAAGATCCC TGAATATGGC  1500
CTACTGGGGT CAAGGAACCT CAGTCACCGT CTCCTCACTA AGCGCAGATG ACGCAAAGAA  1560
AGACGCAGCT AAAAAAGACG ATGCCAAAAA GGATGACGCC AAGAAAGATC TTGACATTGT  1620
GATGTCACAG TCTCCATCCT CCCTACCTGT GTCAGTTGGC GAGAAGGTTA CTTTGAGCTG  1680
CAAGTCCAGT CAGAGCCTTT TATATAGTGG TAATCAAAAG AACTACTTGG CCTGGTACCA  1740
GCAGAAACCA GGGCAGTCTC CTAAACTGCT GATTTACTGG GCATCCGCTA GGGAATCTGG  1800
```

GGTCCCTGAT CGCTTCACAG GCAGTGGATC TGGGACAGAT TTCACTCTCT CCATCAGCAG 1860

TGTGAAGACT GAAGACCTGG CAGTTTATTA CTGTCAGCAG TATTATAGCT ATCCCCTCAC 1920

GTTCGGTGCT GGGACCAAGC TGGTGCTTAA GTAAAAAGCT AGCGATGAAT CCGTCAAAAC 1980

ATCATCTTAC ATAAAGTCAC TTGGTGATCA AGCTCATATC ATTGTCCGGC AATGGTGTGG 2040

GCTTTTTTTG TTTTCTATCT TTAAAGATCA TGTGAAGAAA AACGGGAAAA TCGGTCTGCG 2100

GGAAAGGACC GGGTTTTTGT CGAAATCATA GGCGAATGGG TTGGATTGTG ACAAAATTCG 2160

GATCC 2165

(2) INFORMATION FOR SEQ ID NO:9:

[0126]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 553 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: Protein
(B) LOCATION: 23

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu Leu Ala
        -20                 -15                 -10

Ala Gln Pro Ala Met Ala Asp Ile Val Met Ser Gln Ser Pro Ser Ser
    -5                  1               5                   10

Leu Pro Val Ser Val Gly Glu Lys Val Thr Leu Ser Cys Lys Ser Ser
                15                  20                  25

Gln Ser Leu Leu Tyr Ser Gly Asn Gln Lys Asn Tyr Leu Ala Trp Tyr
            30                  35                  40

Gln Gln Lys Pro Gly Gln Ser Pro Lys Leu Leu Ile Tyr Trp Ala Ser
        45                  50                  55

Ala Arg Glu Ser Gly Val Pro Asp Arg Phe Thr Gly Ser Gly Ser Gly
    60                  65                  70

Thr Asp Phe Thr Leu Ser Ile Ser Ser Val Lys Thr Glu Asp Leu Ala
75                  80                  85                  90
```

```
Val Tyr Tyr Cys Gln Gln Tyr Tyr Ser Tyr Pro Leu Thr Phe Gly Ala
                95                  100                 105

Gly Thr Lys Leu Val Leu Lys Leu Ser Ala Asp Asp Ala Lys Lys Asp
            110                 115                 120

Ala Ala Lys Lys Asp Asp Ala Lys Lys Asp Asp Ala Lys Lys Asp Leu
        125                 130                 135

Glu Val Gln Leu Gln Gln Ser Asp Ala Glu Leu Val Lys Pro Gly Ala
    140                 145                 150

Ser Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Asp His
155                 160                 165                 170

Ala Ile His Trp Val Lys Gln Asn Pro Glu Gln Gly Leu Glu Trp Ile
                175                 180                 185

Gly Tyr Phe Ser Pro Gly Asn Asp Asp Phe Lys Tyr Asn Glu Arg Phe
            190                 195                 200

Lys Gly Lys Ala Thr Leu Thr Ala Asp Lys Ser Ser Ser Thr Ala Tyr
        205                 210                 215

Val Gln Leu Asn Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys
    220                 225                 230

Thr Arg Ser Leu Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser Val Thr
235                 240                 245                 250

Val Ser Ser Leu Ser Ala Asp Asp Ala Lys Lys Asp Ala Ala Lys Lys
                255                 260                 265

Asp Asp Ala Lys Lys Asp Asp Ala Lys Lys Asp Leu Glu Val Gln Leu
            270                 275                 280

Gln Gln Ser Asp Ala Glu Leu Val Lys Pro Gly Ala Ser Val Lys Ile
        285                 290                 295

Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Asp His Ala Ile His Trp
    300                 305                 310

Val Lys Gln Asn Pro Glu Gln Gly Leu Glu Trp Ile Gly Tyr Phe Ser
315                 320                 325                 330

Pro Gly Asn Asp Asp Phe Lys Tyr Asn Glu Arg Phe Lys Gly Lys Ala
                335                 340                 345

Thr Leu Thr Ala Asp Lys Ser Ser Ser Thr Ala Tyr Val Gln Leu Asn
            350                 355                 360
```

```
Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys Thr Arg Ser Leu
        365                 370                 375

Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser Val Thr Val Ser Ser Leu
    380                 385                 390

Ser Ala Asp Asp Ala Lys Lys Asp Ala Ala Lys Lys Asp Asp Ala Lys
395                 400                 405                 410

Lys Asp Asp Ala Lys Lys Asp Leu Asp Ile Val Met Ser Gln Ser Pro
                415                 420                 425

Ser Ser Leu Pro Val Ser Val Gly Glu Lys Val Thr Leu Ser Cys Lys
            430                 435                 440

Ser Ser Gln Ser Leu Leu Tyr Ser Gly Asn Gln Lys Asn Tyr Leu Ala
        445                 450                 455

Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Lys Leu Leu Ile Tyr Trp
    460                 465                 470

Ala Ser Ala Arg Glu Ser Gly Val Pro Asp Arg Phe Thr Gly Ser Gly
475                 480                 485                 490

Ser Gly Thr Asp Phe Thr Leu Ser Ile Ser Ser Val Lys Thr Glu Asp
                495                 500                 505

Leu Ala Val Tyr Tyr Cys Gln Gln Tyr Tyr Ser Tyr Pro Leu Thr Phe
            510                 515                 520

Gly Ala Gly Thr Lys Leu Val Leu Lys
        525                 530
```

(2) INFORMATION FOR SEQ ID NO:10:

[0127]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
TAAACTCGAG GTTCAGTTGC AGCAG                                     25
```

(2) INFORMATION FOR SEQ ID NO:11:

[0128]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 44 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TAAAGCTAGC ACCAAGCGCT TAGTGAGGAG ACGGTGACTG AGGT 44

(2) INFORMATION FOR SEQ ID NO:12:

[0129]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TCGTCCGATT AGGCAAGCTT A 21

(2) INFORMATION FOR SEQ ID NO:13:

[0130]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GATGATTTTA AATACAATGA G 21

(2) INFORMATION FOR SEQ ID NO:14:

[0131]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 98 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TAAATGCGCA GATGACGCAA AGAAAGACGC AGCTAAAAAA GACGATGCCA AAAAGGATGA 60

CGCCAAGAAA GATCTTGAGG TTCAGTTGCA GCAGTCTG 98

(2) INFORMATION FOR SEQ ID NO:15:

[0132]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TGACTTTATG TAAGATGATG T 21

(2) INFORMATION FOR SEQ ID NO:16:

[0133]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 99 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TAAATGCGCA GATGACGCAA AGAAAGACGC AGCTAAAAAA GACGATGCCA AAAAGGATGA 60

CGCCAAGAAA GATCTTGACA TTGTGATGTC ACAGTCTCC 99

(2) INFORMATION FOR SEQ ID NO:17:

[0134]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TAAAGCTAGC TTTTTACTTA AGCACCAGCT TGGTCCC 37

(2) INFORMATION FOR SEQ ID NO:18:

[0135]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CTGCTGGTAC CAGGCCAAG 19

(2) INFORMATION FOR SEQ ID NO:19:

[0136]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID 10:19:

TAAAGCTAGC ACCAAGCGCT TAGTTTCAGC ACCAGCTTGG TCCCAG 46

(2) INFORMATION FOR SEQ ID NO:20:

[0137]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TTGATCACCA AGTGACTTTA TG 22

(2) INFORMATION FOR SEQ ID NO:21:

[0138]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 98 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

TAAGCGCTGA TGATGCTAAG AAGGACGCCG CAAAAAAGGA CGACGCAAAA AAAGATGATG 60

CAAAAAAGGA TCTGGAGGTT CAGTTGCAGC AGTCTGAC 98

(2) INFORMATION FOR SEQ ID NO:22:

[0139]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TTGTGCTAGC TTTTTATGAG GAGACGGTGA CTGAGGTT 38

(2) INFORMATION FOR SEQ ID NO:23:

[0140]

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CAGCAGTATT ATAGCTAT 18

Claims

1. A multivalent single chain antibody which comprises two or more single chain antibody fragments, each single chain antibody fragment specifically binding an antigen, wherein the single chain antibody fragments are covalently linked by a first peptide linker which contains an amino acid sequence of

   Leu Ser Ala Asp Asp Ala Lys Lys Asp Ala Ala Lys Lys Asp
   Asp Ala Lys Lys Asp Asp Ala Lys Lys Asp Leu

   and each single chain antibody fragment comprises

(a) a first polypeptide comprising a light chain variable domain;
(b) a second polypeptide comprising a heavy chain variable domain; and
(c) a second peptide linker linking the first and second polypeptides into a functional binding moiety.

2. The multivalent single chain antibody of claim 1, wherein the light chain variable region and the heavy chain variable region are obtained from antibodies against tumor-associated glycoprotein 72 antigen (TAG-72).

3. The multivalent single chain antibody of claim 1, wherein the light chain variable region has an amino acid sequence of

Asp Ile Val Met Ser Gln Ser Pro Ser Ser Leu Pro Val Ser
Val Gly Glu Lys Val Thr Leu Ser Cys Lys Ser Ser Gln Ser
Leu Leu Tyr Ser Gly Asn Gln Lys Asn Tyr Leu Ala Trp Tyr
Gln Gln Lys Pro Gly Gln Ser Pro Lys Leu Leu Ile Tyr Trp
Ala Ser Ala Arg Glu Ser Gly Val Pro Asp Arg Phe Thr Gly
Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser Ile Ser Ser Val
Lys Thr Glu Asp Leu Ala Val Tyr Tyr Cys Gln Gln Tyr Tyr
Ser Tyr Pro Leu Thr Phe Gly Ala Gly Thr Lys Leu Val Leu
Lys

and the heavy chain variable region has an amino acid sequence of

Glu Val Gln Leu Gln Gln Ser Asp Ala Glu Leu Val Lys Pro
Gly Ala Ser Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr
Phe Thr Asp His Ala Ile His Trp Val Lys Gln Asn Pro Glu
Gln Gly Leu Glu Trp Ile Gly Tyr Phe Ser Pro Gly Asn Asp
Asp Phe Lys Tyr Asn Glu Arg Phe Lys Gly Lys Ala Thr Leu
Thr Ala Asp Lys Ser Ser Ser Thr Ala Tyr Val Gln Leu Asn
Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys Thr Arg
Ser Leu Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser Val Thr
Val Ser Ser.

4. The multivalent single chain antibody of claim 1, wherein the first peptide linker has an amino acid sequence having 25 to 30 amino acid residues.

5. The multivalent single chain antibody of claim 1, wherein the second peptide linker has an amino acid sequence having from 10 to 30 amino acid residues.

6. The multivalent single chain antibody of claim 1, wherein the first and second peptide linkers have substantially the same amino acid sequence.

7. The multivalent single chain antibody of claim 6 wherein the second peptide linker has an amino acid sequence identical to that of the first peptide linker.

8. A DNA sequence which codes for a multivalent single chain antibody, the multivalent single chain antibody comprising two or more single chain antibody fragments, each fragment having affinity for an antigen, wherein the fragments are covalently linked by a first peptide linker which contains an amino acid sequence of

Leu Ser Ala Asp Asp Ala Lys Lys Asp Ala Ala Lys Lys Asp Asp Ala Lys
Lys Asp Asp Ala Lys Lys Asp Leu

and each fragment comprising

(a) a first polypeptide comprising a light chain variable domain;
(b) a second polypeptide comprising a heavy chain variable domain; and
(c) a second peptide linker linking the first and second polypeptides into a functional binding moiety.

9. The DNA sequence of claim 8, which codes for a multivalent single chain antibody, wherein the light chain variable region and the heavy chain variable region are obtained from antibodies against tumor-associated glycoprotein 72 antigen (TAG-72).

10. The DNA sequence of claim 8, wherein the sequence coding for the first polypeptide is substantially homologous to the sequence:

GAC ATT GTG ATG TCA CAG TCT CCA TCC TCC CTA CCT GTG TCA
GTT GGC GAG AAG GTT ACT TTG AGC TGC AAG TCC AGT CAG
AGC CTT TTA TAT AGT GGT AAT CAA AAG AAC TAC TTG GCC
TGG TAC CAG CAG AAA CCA GGG CAG TCT CCT AAA CTG CTG
ATT TAC TGG GCA TCC GCT AGG GAA TCT GGG GTC CCT GAT
CGC TTC ACA GGC AGT GGA TCT GGG ACA GAT TTC ACT CTC
TCC ATC AGC AGT GTG AAG ACT GAA GAC CTG GCA GTT TAT
TAC TGT CAG CAG TAT TAT AGC TAT CCC CTC ACG TTC GGT GCT
GGG ACC AAG CTG GTG CTG AAG

and the first polypeptide retains the characteristic of functional binding to TAG-72
and wherein the sequence coding for the second polypeptide is substantially homologous to the sequence:

GAG GTT CAG TTG CAG CAG TCT GAC GCT GAG TTG GTG AAA
CCT GGG GCT TCA GTG AAG ATT TCC TGC AAG GCT TCT GGC
TAC ACC TTC ACT GAC CAT GCA ATT CAC TGG GTG AAA CAG
AAC CCT GAA CAG GGC CTG GAA TGG ATT GGA TAT TTT TCT
CCC GGA AAT GAT GAT TTT AAA TAC AAT GAG AGG TTC AAG
GGC AAG GCC ACA CTG ACT GCA GAC AAA TCC TCC AGC ACT
GCC TAC GTG CAG CTC AAC AGC CTG ACA TCT GAG GAT TCT
GCA GTG TAT TTC TGT ACA AGA TCC CTG AAT ATG GCC TAC
TGG GGT CAA GGA ACC TCA GTC ACC GTC TCC TCA.

and the second polypeptide retains the characteristic of functional binding to TAG-72.

Patentansprüche

1. Multivalenter einzelkettiger Antikörper, welcher zwei oder mehr einzelkettige Antikörperfragmente umfaßt, wobei jedes einzelkettige Antikörperfragment an ein Antigen spezifisch bindet, worin die einzelkettigen Antikörperfragmente durch einen ersten Peptidlinker kovalent verknüpft sind, welcher eine Aminosäuresequenz von

Leu Ser Ala Asp Asp Ala Lys Lys Asp Ala Ala Lys Lys Asp
Asp Ala Lys Lys Asp Asp Ala Lys Lys Asp Leu

enthält, und jedes einzelkettige Antikörperfragment

(a) ein erstes Polypeptid, umfassend eine variable Domäne einer leichten Kette;
(b) ein zweites Polypeptid, umfassend eine variable Domäne einer schweren Kette; und

(c) einen zweiten Peptidlinker, welcher das erste und das zweite Polypeptid zu einer funktionellen Bindungsgruppe verknüpft,

umfaßt.

2. Multivalenter einzelkettiger Antikörper nach Anspruch 1, worin die variable Region der leichten Kette und die variable Region der schweren Kette aus Antikörpern gegen Tumor-assoziiertes Glykoprotein 72-Antigen (TAG-72) erhalten sind.

3. Multivalenter einzelkettiger Antikörper nach Anspruch 1, worin die variable Region der leichten Kette eine Aminosäuresequenz von

```
Asp Ile Val Met Ser Gln Ser Pro Ser Ser Leu Pro Val Ser
Val Gly Glu Lys Val Thr Leu Ser Cys Lys Ser Ser Gln Ser
Leu Leu Tyr Ser Gly Asn Gln Lys Asn Tyr Leu Ala Trp Tyr
Gln Gln Lys Pro Gly Gln Ser Pro Lys Leu Leu Ile Tyr Trp
Ala Ser Ala Arg Glu Ser Gly Val Pro Asp Arg Phe Thr Gly
Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser Ile Ser Ser Val
Lys Thr Glu Asp Leu Ala Val Tyr Tyr Cys Gln Gln Tyr Tyr
Ser Tyr Pro Leu Thr Phe Gly Ala Gly Thr Lys Leu Val Leu
Lys
```

und die variable Region der schweren Kette eine Aminosäuresequenz von

```
Glu Val Gln Leu Gln Gln Ser Asp Ala Glu Leu Val Lys Pro
Gly Ala Ser Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr
Phe Thr Asp His Ala Ile His Trp Val Lys Gln Asn Pro Glu
Gln Gly Leu Glu Trp Ile Gly Tyr Phe Ser Pro Gly Asn Asp
Asp Phe Lys Tyr Asn Glu Arg Phe Lys Gly Lys Ala Thr Leu
Thr Ala Asp Lys Ser Ser Ser Thr Ala Tyr Val Gln Leu Asn
Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys Thr Arg
Ser Leu Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser Val Thr
Val Ser Ser.
```

aufweist.

4. Multivalenter einzelkettiger Antikörper nach Anspruch 1, worin der erste Peptidlinker eine Aminosäuresequenz mit 25 bis 30 Aminosäureresten aufweist.

5. Multivalenter einzelkettiger Antikörper nach Anspruch 1, worin der zweite Peptidlinker eine Aminosäuresequenz mit von 10 bis 30 Aminosäureresten aufweist.

6. Multivalenter einzelkettiger Antikörper nach Anspruch 1, worin der erste und der zweite Peptidlinker im wesentlichen dieselbe Aminosäuresequenz aufweisen.

7. Multivalenter einzelkettiger Antikörper nach Anspruch 6, worin der zweite Peptidlinker eine Aminosäuresequenz aufweist, die mit derjenigen des ersten Peptidlinkers identisch ist.

8. DNA-Sequenz, welche für einen multivalenten einzelkettigen Antikörper codiert, wobei der multivalente einzelkettige Antikörper zwei oder mehr einzelkettige Antikörperfragmente umfaßt, wobei jedes Fragment eine Affinität für ein Antigen besitzt, worin die Fragmente durch einen ersten Polypeptidlinker kovalent verknüpft sind, welcher eine Aminosäuresequenz von

Leu Ser Ala Asp Asp Ala Lys Lys Asp Ala Ala Lys Lys Asp Asp Ala Lys
Lys Asp Asp Ala Lys Lys Asp Leu

enthält,
und jedes Fragment

(a) ein erstes Polypeptid, umfassend eine variable Domäne einer leichten Kette;
(b) ein zweites Polypeptid, umfassend eine variable Domäne einer schweren Kette; und
(c) einen zweiten Peptidlinker, welcher das erste und das zweite Polypeptid zu einer funktionellen Bindungsgruppe verknüpft,

umfaßt.

9. DNA-Sequenz nach Anspruch 8, welche für einen multivalenten einzelkettigen Antikörper codiert, worin die variable Region der leichten Kette und die variable Region der schweren Kette aus Antikörpern gegen Tumor-assoziiertes Glykoprotein 72-Antigen (TAG-72) erhalten sind.

10. DNA-Sequenz nach Anspruch 8, worin die für das erste Polypeptid codierende Sequenz im wesentlichen homolog zu der Sequenz:

GAC ATT GTG ATG TCA CAG TCT CCA TCC TCC CTA CCT GTG TCA
GTT GGC GAG AAG GTT ACT TTG AGC TGC AAG TCC AGT CAG
AGC CTT TTA TAT AGT GGT AAT CAA AAG AAC TAC TTG GCC
TGG TAC CAG CAG AAA CCA GGG CAG TCT CCT AAA CTG CTG
ATT TAC TGG GCA TCC GCT AGG GAA TCT GGG GTC CCT GAT
CGC TTC ACA GGC AGT GGA TCT GGG ACA GAT TTC ACT CTC
TCC ATC AGC AGT GTG AAG ACT GAA GAC CTG GCA GTT TAT
TAC TGT CAG CAG TAT TAT AGC TAT CCC CTC ACG TTC GGT GCT
GGG ACC AAG CTG GTG CTG AAG

ist und das erste Polypeptid die Eigenschaft der funktionellen Bindung an TAG-72 beibehält, und worin die für das zweite Polypeptid codierende Sequenz im wesentlichen homolog zu der Sequenz

GAG GTT CAG TTG CAG CAG TCT GAC GCT GAG TTG GTG AAA
CCT GGG GCT TCA GTG AAG ATT TCC TGC AAG GCT TCT GGC
TAC ACC TTC ACT GAC CAT GCA ATT CAC TGG GTG AAA CAG
AAC CCT GAA CAG GGC CTG GAA TGG ATT GGA TAT TTT TCT
CCC GGA AAT GAT GAT TTT AAA TAC AAT GAG AGG TTC AAG
GGC AAG GCC ACA CTG ACT GCA GAC AAA TCC TCC AGC ACT
GCC TAC GTG CAG CTC AAC AGC CTG ACA TCT GAG GAT TCT
GCA GTG TAT TTC TGT ACA AGA TCC CTG AAT ATG GCC TAC
TGG GGT CAA GGA ACC TCA GTC ACC GTC TCC TCA.

ist und das zweite Polypeptid die Eigenschaft einer funktionellen Bindung an TAG-72 beibehält.

Revendications

1. Anticorps monocaténaire multivalent qui comporte deux fragments monocaténaires d'anticorps ou plus, chaque fragment monocaténaire d'anticorps fixant spécifiquement un antigène, dans lequel anticorps les fragments monocaténaires d'anticorps sont reliés de manière covalente par un premier raccord peptidique qui contient la séquence d'acides aminés suivante :

   Leu Ser Ala Asp Asp Ala Lys Lys Asp Ala Ala Lys Lys Asp
   Asp Ala Lys Lys Asp Asp Ala Lys Lys Asp Leu

   et chaque fragment monocaténaire d'anticorps comporte

   a) un premier polypeptide constituant un domaine variable de chaîne légère ;
   b) un deuxième polypeptide constituant un domaine variable de chaîne lourde ; et
   c) un deuxième raccord peptidique reliant ces premier et deuxième polypeptides en une entité fixatrice fonctionnelle.

2. Anticorps monocaténaire multivalent, conforme à la revendication 1, dans lequel le domaine variable de chaîne légère et le domaine variable de chaîne lourde ont été obtenus à partir d'anticorps dirigés contre l'antigène qu'est la glycoprotéine 72 associée à des tumeurs (antigène TAG-72).

3. Anticorps monocaténaire multivalent, conforme à la revendication 1, dans lequel le domaine variable de chaîne légère présente la séquence suivante d'acides aminés :

   Asp Ile Val Met Ser Gln Ser Pro Ser Ser Leu Pro Val Ser
   Val Gly Glu Lys Val Thr Leu Ser Cys Lys Ser Ser Gln Ser
   Leu Leu Tyr Ser Gly Asn Gln Lys Asn Tyr Leu Ala Trp Tyr
   Gln Gln Lys Pro Gly Gln Ser Pro Lys Leu Leu Ile Tyr Trp
   Ala Ser Ala Arg Glu Ser Gly Val Pro Asp Arg Phe Thr Gly
   Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser Ile Ser Ser Val
   Lys Thr Glu Asp Leu Ala Val Tyr Tyr Cys Gln Gln Tyr Tyr
   Ser Tyr Pro Leu Thr Phe Gly Ala Gly Thr Lys Leu Val Leu
   Lys

   et le domaine variable de chaîne lourde présente la séquence suivante d'acides aminés :

   Glu Val Gln Leu Gln Gln Ser Asp Ala Glu Leu Val Lys Pro
   Gly Ala Ser Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr
   Phe Thr Asp His Ala Ile His Trp Val Lys Gln Asn Pro Glu
   Gln Gly Leu Glu Trp Ile Gly Tyr Phe Ser Pro Gly Asn Asp
   Asp Phe Lys Tyr Asn Glu Arg Phe Lys Gly Lys Ala Thr Leu
   Thr Ala Asp Lys Ser Ser Ser Thr Ala Tyr Val Gln Leu Asn
   Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys Thr Arg
   Ser Leu Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser Val Thr
   Val Ser Ser.

4. Anticorps monocaténaire multivalent, conforme à la revendication 1, dans lequel la séquence d'acides aminés du premier raccord peptidique comporte de 25 à 30 résidus d'acides aminés.

5. Anticorps monocaténaire multivalent, conforme à la revendication 1, dans lequel la séquence d'acides aminés du deuxième raccord peptidique comporte de 10 à 30 résidus d'acides aminés.

6. Anticorps monocaténaire multivalent, conforme à la revendication 1, dans lequel les premier et deuxième raccords

peptidiques ont pratiquement la même séquence d'acides aminés.

7. Anticorps monocaténaire multivalent, conforme à la revendication 1, dans lequel la séquence d'acides aminés du deuxième raccord peptidique est identique à celle du premier raccord peptidique.

8. Séquence d'ADN qui code un anticorps monocaténaire multivalent, lequel anticorps monocaténaire multivalent comporte deux fragments monocaténaires d'anticorps ou plus, chacun de ces fragments présentant une certaine affinité pour un antigène, et ces fragments étant reliés de manière covalente par un premier raccord peptidique qui contient la séquence d'acides aminés suivante :

Leu Ser Ala Asp Asp Ala Lys Lys Asp Ala Ala Lys Lys Asp
Asp Ala Lys Lys Asp Asp Ala Lys Lys Asp Leu

et chaque fragment comportant

a) un premier polypeptide constituant un domaine variable de chaîne légère ;
b) un deuxième polypeptide constituant un domaine variable de chaîne lourde ; et
c) un deuxième raccord peptidique reliant ces premier et deuxième polypeptides en une entité fixatrice fonctionnelle.

9. Séquence d'ADN conforme à la revendication 8, qui code un anticorps monocaténaires multivalent dans lequel le domaine variable de chaîne légère et le domaine variable de chaîne lourde ont été obtenus à partir d'anticorps dirigés contre l'antigène qu'est la glycoprotéine 72 associée à des tumeurs (antigène TAG-72).

10. Séquence d'ADN conforme à la revendication 8, dans laquelle la séquence codant le premier polypeptide est pratiquement homologue de la séquence suivante :

GAC ATT GTG ATG TCA CAG TCT CCA TCC TCC CTA CCT GTG TCA
GTT GGC GAG AAG GTT ACT TTG AGC TGC AAG TCC AGT CAG
AGC CTT TTA TAT AGT GGT AAT CAA AAG AAC TAC TTG GCC
TGG TAC CAG CAG AAA CCA GGG CAG TCT CCT AAA CTG CTG
ATT TAC TGG GCA TCC GCT AGG GAA TCT GGG GTC CCT GAT
CGC TTC ACA GGC AGT GGA TCT GGG ACA GAT TTC ACT CTC
TCC ATC AGC AGT GTG AAG ACT GAA GAC CTG GCA GTT TAT
TAC TGT CAG CAG TAT TAT AGC TAT CCC CTC ACG TTC GGT GCT
GGG ACC AAG CTG GTG CTG AAG

le premier polypeptide conservant la propriété de liaison fonctionnelle à l'antigène TAG-72,
et la séquence codant le deuxième polypeptide est pratiquement homologue de la séquence suivante :

GAG GTT CAG TTG CAG CAG TCT GAC GCT GAG TTG GTG AAA
CCT GGG GCT TCA GTG AAG ATT TCC TGC AAG GCT TCT GGC
TAC ACC TTC ACT GAC CAT GCA ATT CAC TGG GTG AAA CAG
AAC CCT GAA CAG GGC CTG GAA TGG ATT GGA TAT TTT TCT
CCC GGA AAT GAT GAT TTT AAA TAC AAT GAG AGG TTC AAG
GGC AAG GCC ACA CTG ACT GCA GAC AAA TCC TCC AGC ACT
GCC TAC GTG CAG CTC AAC AGC CTG ACA TCT GAG GAT TCT
GCA GTG TAT TTC TGT ACA AGA TCC CTG AAT ATG GCC TAC
TGG GGT CAA GGA ACC TCA GTC ACC GTC TCC TCA.

le deuxième polypeptide conservant la propriété de liaison fonctionnelle à l'antigène TAG-72.

FIG. 1

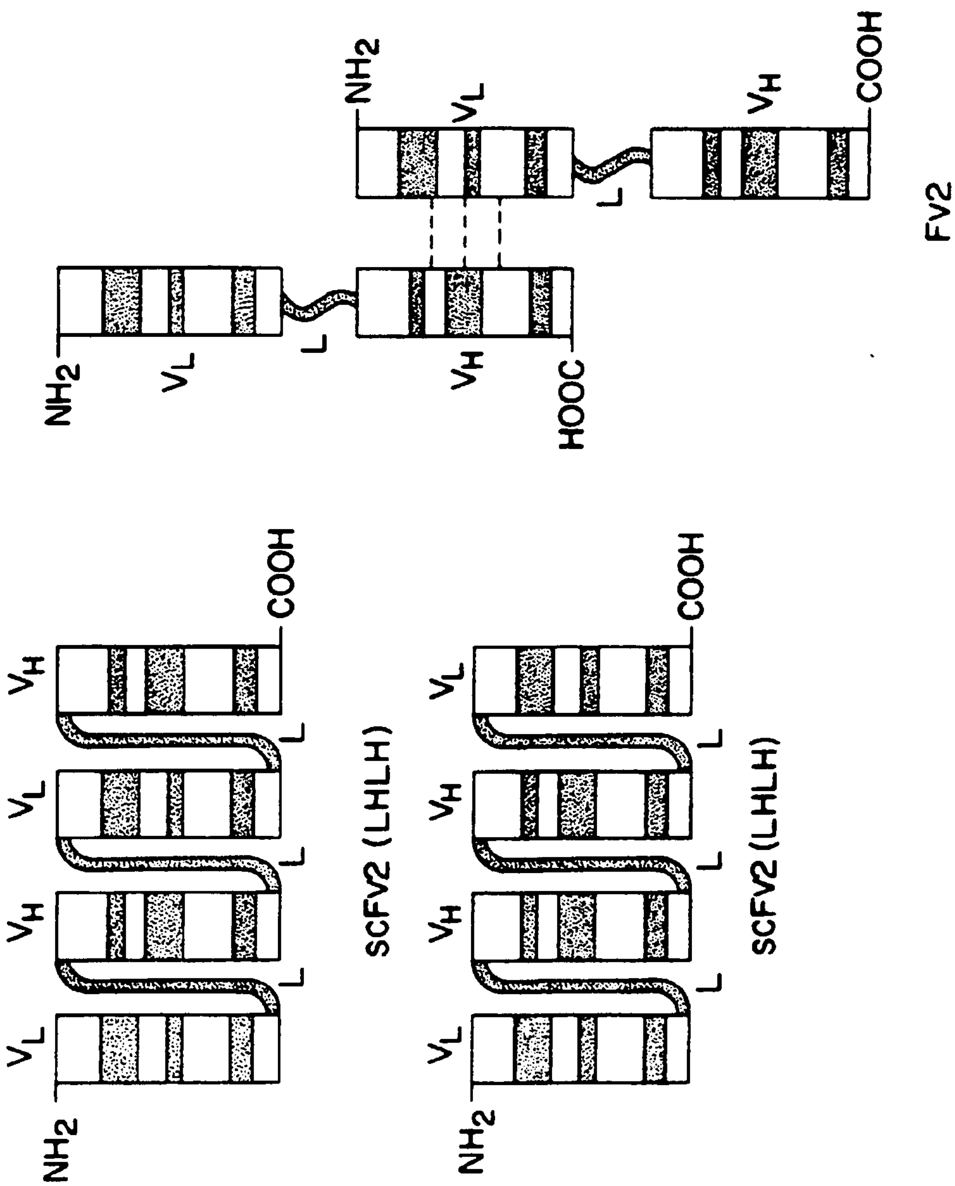

NH2
VL
L
VH
L
VL
L
VH
COOH
scFV2 (LHLH)
NH2
VL
L
VH
L
VH
L
VL
COOH
scFV2 (LHLH)
NH2
VL
L
VH
HOOC
NH2
VL
L
VH
COOH
FV2

FIG. 2

GAC ATT GTG ATG TCA CAG TCT CCA TCC TCC CTA CCT GTG TCA
GTT GGC GAG AAG GTT ACT TTG AGC TGC AAG TCC AGT CAG AGC
CTT TTA TAT AGT GGT AAT CAA AAG AAC TAC TTG GCC TGG TAC
CAG CAG AAA CCA GGG CAG TCT CCT AAA CTG CTG ATT TAC TGG
GCA TCC GCT AGG GAA TCT GGG GTC CCT GAT CGC TTC ACA GGC
AGT GGA TCT GGG ACA GAT TTC ACT CTC TCC ATC AGC AGT GTG
AAG ACT GAA GAC CTG GCA GTT TAT TAC TGT CAG CAG TAT TAT
AGC TAT CCC CTC ACG TTC GGT GCT GGG ACC AAG CTG GTG CTG
AAG

FIG. 3

Asp Ile Val Met Ser Gln Ser Pro Ser Ser Leu Pro Val Ser Val
Gly Glu Lys Val Thr Leu Ser Cys Lys Ser Ser Gln Ser Leu Leu
Tyr Ser Gly Asn Gln Lys Asn Tyr Leu Ala Trp Tyr Gln Gln Lys
Pro Gly Gln Ser Pro Lys Leu Leu Ile Tyr Trp Ala Ser Ala Arg
Glu Ser Gly Val Pro Asp Arg Phe Thr Gly Ser Gly Ser Gly Thr
Asp Phe Thr Leu Ser Ile Ser Ser Val Lys Thr Glu Asp Leu Ala
Val Tyr Tyr Cys Gln Gln Tyr Tyr Ser Tyr Pro Leu Thr Phe Gly
Ala Gly Thr Lys Leu Val Leu Lys

FIG. 4

GAG GTT CAG TTG CAG CAG TCT GAC GCT GAG TTG GTG AAA CCT
GGG GCT TCA GTG AAG ATT TCC TGC AAG GCT TCT GGC TAC ACC
TTC ACT GAC CAT GCA ATT CAC TGG GTG AAA CAG AAC CCT GAA
CAG GGC CTG GAA TGG ATT GGA TAT TTT TCT CCC GGA AAT GAT
GAT TTT AAA TAC AAT GAG AGG TTC AAG GGC AAG GCC ACA CTG
ACT GCA GAC AAA TCC TCC AGC ACT GCC TAC GTG CAG CTC AAC
AGC CTG ACA TCT GAG GAT TCT GCA GTG TAT TTC TGT ACA AGA
TCC CTG AAT ATG GCC TAC TGG GGT CAA GGA ACC TCA GTC ACC
GTC TCC TCA

FIG. 5

Glu Val Gln Leu Gln Gln Ser Asp Ala Glu Leu Val Lys Pro Gly
Ala Ser Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr
Asp His Ala Ile His Trp Val Lys Gln Asn Pro Glu Gln Gly Leu
Glu Trp Ile Gly Tyr Phe Ser Pro Gly Asn Asp Asp Phe Lys Tyr
Asn Glu Arg Phe Lys Gly Lys Ala Thr Leu Thr Ala Asp Lys Ser
Ser Ser Thr Ala Tyr Val Gln Leu Asn Ser Leu Thr Ser Glu Asp
Ser Ala Val Tyr Phe Cys Thr Arg Ser Leu Asn Met Ala Tyr Trp
Gly Gln Gly Thr Ser Val Thr Val Ser Ser

FIG. 6A

DNA AND AMINO ACID SEQUENCE OF CC49 V_L-L-V_H-L-V_L-L-V_H

```
                                   Cla I   EcoR I
5'-C TCA TGT TTG ACA GCT TAT CAT CGA TGA ATT CCA TCA CTT CCC TCC     46

 GTT CAT TTG TCC CCG GTG GAA ACG AGG TCA TCA TTT CCT TCC GAA AAA     94

 ACG GTT GCA TTT AAA TCT TAC ATA TAT AAT ACT TTC AAA GAC TAC ATT    142

 TGT AAG ATT TGA TGT TTG AGT CGG CTG AAA GAT CGT ACG TAC CAA TTA    190

                         PENPR1- AAC ACT GTA GGG ATA GTG GAA
 TTG TTT CGT GAT TGT TCA AGC CAT AAC ACT GTA GGG ATA GTG GAA AGA    238

 GTG CTT CAT CTG GTT ACG ATC AAT CAA ATA TTC AAA CGG AGG GAG ACG    286
                             PENPR2- TAT AAG TTT GCC TCC CTC TG
         -22
         Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu
 ATT TTG ATG AAA TAC CTA TTG CCT ACG GCA GCC GCT GGA TTG TTA TTA    334

                     Nco I       VL
 Leu Ala Ala Gln Pro Ala Met Ala Asp Ile Val Met Ser Gln Ser Pro
 CTC GCT GCC CAA CCA GCC ATG GCC GAC ATT GTG ATG TCA CAG TCT CCA    382

      10                                     20
 Ser Ser Leu Pro Val Ser Val Gly Glu Lys Val Thr Leu Ser Cys Lys
 TCC TCC CTA CCT GTG TCA GTT GGC GAG AAG GTT ACT TTG AGC TGC AAG    430

                     30                                      40
 Ser Ser Gln Ser Leu Leu Tyr Ser Gly Asn Gln Lys Asn Tyr Leu Ala
 TCC AGT CAG AGC CTT TTA TAT AGT GGT AAT CAA AAG AAC TAC TTG GCC    478
```

FIG. 6B

```
                                          50
Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Lys Leu Leu Ile Tyr Trp
TGG TAC CAG CAG AAA CCA GGG CAG TCT CCT AAA CTG CTG ATT TAC TGG     526

            60                                      70
Ala Ser Ala Arg Glu Ser Gly Val Pro Asp Arg Phe Thr Gly Ser Gly
GCA TCC GCT AGG GAA TCT GGG GTC CCT GAT CGC TTC ACA GGC AGT GGA     574

                             80
Ser Gly Thr Asp Phe Thr Leu Ser Ile Ser Ser Val Lys Thr Glu Asp
TCT GGG ACA GAT TTC ACT CTC TCC ATC AGC AGT GTG AAG ACT GAA GAC     622

    90                                      100
Leu Ala Val Tyr Tyr Cys Gln Gln Tyr Tyr Ser Tyr Pro Leu Thr Phe
CTG GCA GTT TAT TAC TGT CAG CAG TAT TAT AGC TAT CCC CTC ACG TTC     670

                    110         Hind III                    120
Gly Ala Gly Thr Lys Leu Val Leu Lys Leu Ser Ala Asp Asp Ala Lys
GGT GCT GGG ACC AAG CTG GTG CTG AAG CTT AGT GCG GAC GAT GCG AAA     718

                                         130
Lys Asp Ala Ala Lys Lys Asp Asp Ala Lys Lys Asp Asp Ala Lys Lys
AAG GAT GCT GCG AAG AAG GAT GAC GCT AAG AAA GAC GAT GCT AAA AAG     766
```

FIG. 6C

```
        VH
    Xho I   140                                            150
Asp Leu Glu Val Gln Leu Gln Gln Ser Asp Ala Glu Leu Val Lys Pro
GAC CTC GAG GTT CAG TTG CAG CAG TCT GAC GCT GAG TTG GTG AAA CCT     814

                            160
Gly Ala Ser Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr
GGG GCT TCA GTG AAG ATT TCC TGC AAG GCT TCT GGC TAC ACC TTC ACT     862

    170                                     180
Asp His Ala Ile His Trp Val Lys Gln Asn Pro Glu Gln Gly Leu Glu
GAC CAT GCA ATT CAC TGG GTG AAA CAG AAC CCT GAA CAG GGC CTG GAA     910

                                                    200
                        190  CC49VHP- GAT GAT TTT AAA TAC AAT GAG
Trp Ile Gly Tyr Phe Ser Pro Gly Asn Asp Asp Phe Lys Tyr Asn Glu
TGG ATT GGA TAT TTT TCT CCC GGA AAT GAT GAT TTT AAA TAC AAT GAG     958

                                    210
Arg Phe Lys Gly Lys Ala Thr Leu Thr Ala Asp Lys Ser Ser Ser Thr
AGG TTC AAG GGC AAG GCC ACA CTG ACT GCA GAC AAA TCC TCC AGC ACT    1006

            220                                     230
Ala Tyr Val Gln Leu Asn Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr
GCC TAC GTG CAG CTC AAC AGC CTG ACA TCT GAG GAT TCT GCA GTG TAT    1054
```

FIG. 6D

```
                                      240
                   VH49J- G AAT ATG GCC TAC TGG GGT CAA G
Phe Cys Thr Arg Ser Leu Asn Met Ala /Tyr Trp Gly Gln Gly Thr Ser
TTC TGT ACA AGA TCC CTG AAT ATG GCC TAC TGG GGT CAA GGA ACC TCA     1102

    250                                     260
Val Thr Val Ser Ser Leu Ser Ala Asp Asp Ala Lys Lys Asp Ala Ala
GTC ACC GTC TCC TCA CTA AGC GCA GAT GAC GCA AAG AAA GAC GCA GCT     1150

                    270                                 VL  280
Lys Lys Asp Asp Ala Lys Lys Asp Asp Ala Lys Lys Asp Leu Asp Ile
AAA AAA GAC GAT GCC AAA AAG GAT GAC GCC AAG AAA GAT CTT GAC ATT     1198

                                        290
Val Met Ser Gln Ser Pro Ser Ser Leu Pro Val Ser Val Gly Glu Lys
GTG ATG TCA CAG TCT CCA TCC TCC CTA CCT GTG TCA GTT GGC GAG AAG     1246

            300                                     310
Val Thr Leu Ser Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser Gly Asn
GTT ACT TTG AGC TGC AAG TCC AGT CAG AGC CTT TTA TAT AGT GGT AAT     1294

                                320
Gln Lys Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro
CAA AAG AAC TAC TTG GCC TGG TAC CAG CAG AAA CCA GGG CAG TCT CCT     1342
```

FIG. 6E

```
    330                                          340
Lys Leu Leu Ile Tyr Trp Ala Ser Ala Arg Glu Ser Gly Val Pro Asp
AAA CTG CTG ATT TAC TGG GCA TCC GCT AGG GAA TCT GGG GTC CCT GAT     1390

                        350                                     360
Arg Phe Thr Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser Ile Ser
CGC TTC ACA GGC AGT GGA TCT GGG ACA GAT TTC ACT CTC TCC ATC AGC     1438

                                    370
                                     49VLCDR3+- CAG CAG TAT TAT
Ser Val Lys Thr Glu Asp Leu Ala Val Tyr Tyr Cys Gln Gln Tyr Tyr
AGT GTG AAG ACT GAA GAC CTG GCA GTT TAT TAC TGT CAG CAG TAT TAT     1486

AGC TAT     380                                     390
Ser Tyr Pro Leu Thr Phe Gly Ala Gly Thr Lys Leu Val Leu Lys Leu
AGC TAT CCC CTC ACG TTC GGT GCT GGG ACC AAG CTG GTG CTG AAG CTA     1534

Eco47 III                   400
Ser Ala Asp Asp Ala Lys Lys Asp Ala Ala Lys Lys Asp Asp Ala Lys
AGC GCT GAT GAT GCT AAG AAG GAC GCC GCA AAA AAG GAC GAC GCA AAA     1582

    410                             VH          420
Lys Asp Asp Ala Lys Lys Asp Leu Glu Val Gln Leu Gln Gln Ser Asp
AAG GAT GAT GCA AAA AAG GAT CTG GAG GTT CAG TTG CAG CAG TCT GAC     1630

                        430                                     440
Ala Glu Leu Val Lys Pro Gly Ala Ser Val Lys Ile Ser Cys Lys Ala
GCT GAG TTG GTG AAA CCT GGG GCT TCA GTG AAG ATT TCC TGC AAG GCT     1678
```

FIG. 6F

```
                                            450
Ser Gly Tyr Thr Phe Thr Asp His Ala Ile His Trp Val Lys Gln Asn
TCT GGC TAC ACC TTC ACT GAC CAT GCA ATT CAC TGG GTG AAA CAG AAC    1726

            460                                     470
Pro Glu Gln Gly Leu Glu Trp Ile Gly Tyr Phe Ser Pro Gly Asn Asp
CCT GAA CAG GGC CTG GAA TGG ATT GGA TAT TTT TCT CCC GGA AAT GAT    1774

                            480
Asp Phe Lys Tyr Asn Glu Arg Phe Lys Gly Lys Ala Thr Leu Thr Ala
GAT TTT AAA TAC AAT GAG AGG TTC AAG GGC AAG GCC ACA CTG ACT GCA    1822

    490                                         500
Asp Lys Ser Ser Ser Thr Ala Tyr Val Gln Leu Asn Ser Leu Thr Ser
GAC AAA TCC TCC AGC ACT GCC TAC GTG CAG CTC AAC AGC CTG ACA TCT    1870

                    510                                     520
Glu Asp Ser Ala Val Tyr Phe Cys Thr Arg Ser Leu Asn Met Ala Tyr
GAG GAT TCT GCA GTG TAT TTC TGT ACA AGA TCC CTG AAT ATG GCC TAC    1918

                                    530
Trp Gly Gln Gly Thr Ser Val Thr Val Ser Ser ***     Nhe I
TGG GGT CAA GGA ACC TCA GTC ACC GTC TCC TCA TAA AAA GCT AGC GAT    1966
```

FIG. 6G

```
GAA TCC GTC AAA ACA TCA TCT TAC ATA AAG TCA CTT GGT GAT CAA GCT    2014
          SQP1- TGT AGT AGA ATG TAT TTC AGT
                      PENPTSEQ2- G TAT TTC AGT GAA CCA CTA GTT

CAT ATC ATT GTC CGG CAA TGG TGT GGG CTT TTT TTG TTT TCT ATC TTT    2062

AAA GAT CAT GTG AAG AAA AAC GGG AAA ATC GGT CTG CGG GAA AGG ACC    2110

GGG TTT TTG TCG AAA TCA TAG GCG AAT GGG TTG GAT TGT GAC AAA ATT    2158

 BamH I
CGG ATC C-3'                                                       2165
```

FIG. 7A

DNA AND AMINO ACID SEQUENCE OF CC49 V_L-L-V_H-L-V_H-L-V_L

```
                                     Cla I   EooR I
5'-C TCA TGT TTG ACA GCT TAT CAT CGA TGA ATT CCA TCA CTT CCC TCC      46

 GTT CAT TTG TCC CCG GTG GAA ACG AGG TCA TCA TTT CCT TCC GAA AAA      94

 ACG GTT GCA TTT AAA TCT TAC ATA TAT AAT ACT TTC AAA GAC TAC ATT     142

 TGT AAG ATT TGA TGT TTG AGT CGG CTG AAA GAT CGT ACG TAC CAA TTA     190

                         PENPR1- AAC ACT GTA GGG ATA GTG GAA
 TTG TTT CGT GAT TGT TCA AGC CAT AAC ACT GTA GGG ATA GTG GAA AGA     238

 GTG CTT CAT CTG GTT ACG ATC AAT CAA ATA TTC AAA CGG AGG GAG ACG     286
                             PENPR2- TAT AAG TTT GCC TCC CTC TG

         -22
         Met Lys Tyr Leu Leu Pro Thr Ala Ala Ala Gly Leu Leu Leu
 ATT TTG ATG AAA TAC CTA TTG CCT ACG GCA GCC GCT GGA TTG TTA TTA     334

                     Nco I       VL
 Leu Ala Ala Gln Pro Ala Met Ala Asp Ile Val Met Ser Gln Ser Pro
 CTC GCT GCC CAA CCA GCC ATG GCC GAC ATT GTG ATG TCA CAG TCT CCA     382

      10                                     20
 Ser Ser Leu Pro Val Ser Val Gly Glu Lys Val Thr Leu Ser Cys Lys
 TCC TCC CTA CCT GTG TCA GTT GGC GAG AAG GTT ACT TTG AGC TGC AAG     430
```

FIG. 7B

```
                                30                                         40
Ser Ser Gln Ser Leu Leu Tyr Ser Gly Asn Gln Lys Asn Tyr Leu Ala
TCC AGT CAG AGC CTT TTA TAT AGT GGT AAT CAA AAG AAC TAC TTG GCC     478
                                        50
Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Lys Leu Leu Ile Tyr Trp
TGG TAC CAG CAG AAA CCA GGG CAG TCT GCT AAA CTG CTG ATT TAC TGG     526

            60                                         70
Ala Ser Ala Arg Glu Ser Gly Val Pro Asp Arg Phe Thr Gly Ser Gly
GCA TCC GCT AGG GAA TCT GGG GTC CCT GAT CGC TTC ACA GGC AGT GGA     574

                                    80
Ser Gly Thr Asp Phe Thr Leu Ser Ile Ser Ser Val Lys Thr Glu Asp
TCT GGG ACA GAT TTC ACT CTC TCC ATC AGC AGT GTG AAG ACT GAA GAC     622

    90                                              100
Leu Ala Val Tyr Tyr Cys Gln Gln Tyr Tyr Ser Tyr Pro Leu Thr Phe
CTG GCA GTT TAT TAC TGT CAG CAG TAT TAT AGC TAT CCC CTC ACG TTC     670

                    110         Hind III                       120
Gly Ala Gly Thr Lys Leu Val Leu Lys Leu Ser Ala Asp Asp Ala Lys
GGT GCT GGG ACC AAG CTG GTG CTG AAG CTT AGT GCG GAC GAT GCG AAA     718

                                        130
Lys Asp Ala Ala Lys Lys Asp Asp Ala Lys Lys Asp Asp Ala Lys Lys
AAG GAT GCT GCG AAG AAG GAT GAC GCT AAG AAA GAC GAT GCT AAA AAG     766
TTC CTA CGA CGC TTC TTC CTA TMNVL(-)SEQ
```

FIG. 7C

```
         VH
     Xho I   140                                         150
Asp Leu Glu Val Gln Leu Gln Gln Ser Asp Ala Glu Leu Val Lys Pro
GAC CTC GAG GTT CAG TTG CAG CAG TCT GAC GCT GAG TTG GTG AAA CCT    814

                            160
Gly Ala Ser Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr
GGG GCT TCA GTG AAG ATT TCC TGC AAG GCT TCT GGC TAC ACC TTC ACT    862

    170                                     180
Asp His Ala Ile His Trp Val Lys Gln Asn Pro Glu Gln Gly Leu Glu
GAC CAT GCA ATT CAC TGG GTG AAA CAG AAC CCT GAA CAG GGC CTG GAA    910

                                                            200
                    190    CC49VHP- GAT GAT TTT AAA TAC AAT GAG
Trp Ile Gly Tyr Phe Ser Pro Gly Asn Asp Asp Phe Lys Tyr Asn Glu
TGG ATT GGA TAT TTT TCT CCC GGA AAT GAT GAT TTT AAA TAC AAT GAG    958

                                    210
Arg Phe Lys Gly Lys Ala Thr Leu Thr Ala Asp Lys Ser Ser Ser Thr
AGG TTC AAG GGC AAG GCC ACA CTG ACT GCA GAC AAA TCC TCC AGC ACT   1006

        220                                         230
Ala Tyr Val Gln Leu Asn Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr
GCC TAC GTG CAG CTC AAC AGC CTG ACA TCT GAG GAT TCT GCA GTG TAT   1054

                            240
                VH49J- G AAT ATG GCC TAC TGG GGT CAA G
Phe Cys Thr Arg Ser Leu Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser
TTC TGT ACA AGA TCC CTG AAT ATG GCC TAC TGG GGT CAA GGA ACC TCA   1102
```

FIG. 7D

```
 250                                                260
Val Thr Val Ser Ser Leu Ser Ala Asp Asp Ala Lys Lys Asp Ala Ala
GTC ACC GTC TCC TCA CTA AGC GCA GAT GAC GCA AAG AAA GAC GCA GCT   1150

                     270                                VH  280
Lys Lys Asp Asp Ala Lys Lys Asp Asp Ala Lys Lys Asp Leu Glu Val
AAA AAA GAC GAT GCC AAA AAG GAT GAC GCC AAG AAA GAT CTT GAG GTT   1198

                                     290
Gln Leu Gln Gln Ser Asp Ala Glu Leu Val Lys Pro Gly Ala Ser Val
CAG TTG CAG CAG TCT GAC GCT GAG TTG GTG AAA CCT GGG GCT TCA GTG   1246

             300                                     310
Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Asp His Ala Ile
AAG ATT TCC TGC AAG GCT TCT GGC TAC ACC TTC ACT GAC CAT GCA ATT   1294

                                 320
His Trp Val Lys Gln Asn Pro Glu Gln Gly Leu Glu Trp Ile Gly Tyr
CAC TGG GTG AAA CAG AAC CCT GAA CAG GGC CTG GAA TGG ATT GGA TAT   1342

     330                                         340
Phe Ser Pro Gly Asn Asp Asp Phe Lys Tyr Asn Glu Arg Phe Lys Gly
TTT TCT CCC GGA AAT GAT GAT TTT AAA TAC AAT GAG AGG TTC AAG GGC   1390

                     350                                     360
Lys Ala Thr Leu Thr Ala Asp Lys Ser Ser Ser Thr Ala Tyr Val Gln
AAG GCC ACA CTG ACT GCA GAC AAA TCC TCC AGC ACT GCC TAC GTG CAG   1438

                                     370
Leu Asn Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys Thr Arg
CTC AAC AGC CTG ACA TCT GAG GAT TCT GCA GTG TAT TTC TGT ACA AGA   1486
```

FIG. 7F

```
    490                                          500
Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser Ile Ser Ser Val Lys Thr
AGT GGA TCT GGG ACA GAT TTC ACT CTC TCC ATC AGC AGT GTG AAG ACT    1870

                    510                                      520
Glu Asp Leu Ala Val Tyr Tyr Cys Gln Gln Tyr Tyr Ser Tyr Pro Leu
GAA GAC CTG GCA GTT TAT TAC TGT CAG CAG TAT TAT AGC TAT CCC CTC    1918
                                    530
                                    Afl II
Thr Phe Gly Ala Gly Thr Lys Leu Val Leu Lys ***     Nhe I
ACG TTC GGT GCT GGG ACC AAG CTG GTG CTT AAG TAA AAA GCT AGC GAT    1966

GAA TCC GTC AAA ACA TCA TCT TAC ATA AAG TCA CTT GGT GAT CAA GCT    2014
          SQP1- TGT AGT AGA ATG TAT TTC AGT
                     PENPTSEQ2- G TAT TTC AGT GAA CCA CTA GTT

CAT ATC ATT GTC CGG CAA TGG TGT GGG CTT TTT TTG TTT TCT ATC TTT    2062

AAA GAT CAT GTG AAG AAA AAC GGG AAA ATC GGT CTG CGG GAA AGG ACC    2110

GGG TTT TTG TCG AAA TCA TAG GCG AAT GGG TTG GAT TGT GAC AAA ATT    2158

 BamH I
CGG ATC C-3'                                                       2165
```

FIG. 7E

```
                380                                                      390
Ser Leu Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser Val Thr Val Ser
TCC CTG AAT ATG GCC TAC TGG GGT CAA GGA ACC TCA GTC ACC GTC TCC    1534

                                400
Ser Leu Ser Ala Asp Asp Ala Lys Lys Asp Ala Ala Lys Lys Asp Asp
TCA CTA AGC GCA GAT GAC GCA AAG AAA GAC GCA GCT AAA AAA GAC GAT    1582

    410                                         VL  420
Ala Lys Lys Asp Asp Ala Lys Lys Asp Leu Asp Ile Val Met Ser Gln
GCC AAA AAG GAT GAC GCC AAG AAA GAT CTT GAC ATT GTG ATG TCA CAG    1630

                    430                                         440
Ser Pro Ser Ser Leu Pro Val Ser Val Gly Glu Lys Val Thr Leu Ser
TCT CCA TCC TCC CTA CCT GTG TCA GTT GGC GAG AAG GTT ACT TTG AGC    1678

                                        450
Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser Gly Asn Gln Lys Asn Tyr
TGC AAG TCC AGT CAG AGC CTT TTA TAT AGT GGT AAT CAA AAG AAC TAC    1726
                                                   49LFR2(-)- G

                460                                     470
Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Lys Leu Leu Ile
TTG GCC TGG TAC CAG CAG AAA CCA GGG CAG TCT CCT AAA CTG CTG ATT    1774
AAC CGG ACC ATG GTC GTC

                                480
Tyr Trp Ala Ser Ala Arg Glu Ser Gly Val Pro Asp Arg Phe Thr Gly
TAC TGG GCA TCC GCT AGG GAA TCT GGG GTC CCT GAT CGC TTC ACA GGC    1822
```

FIG.8

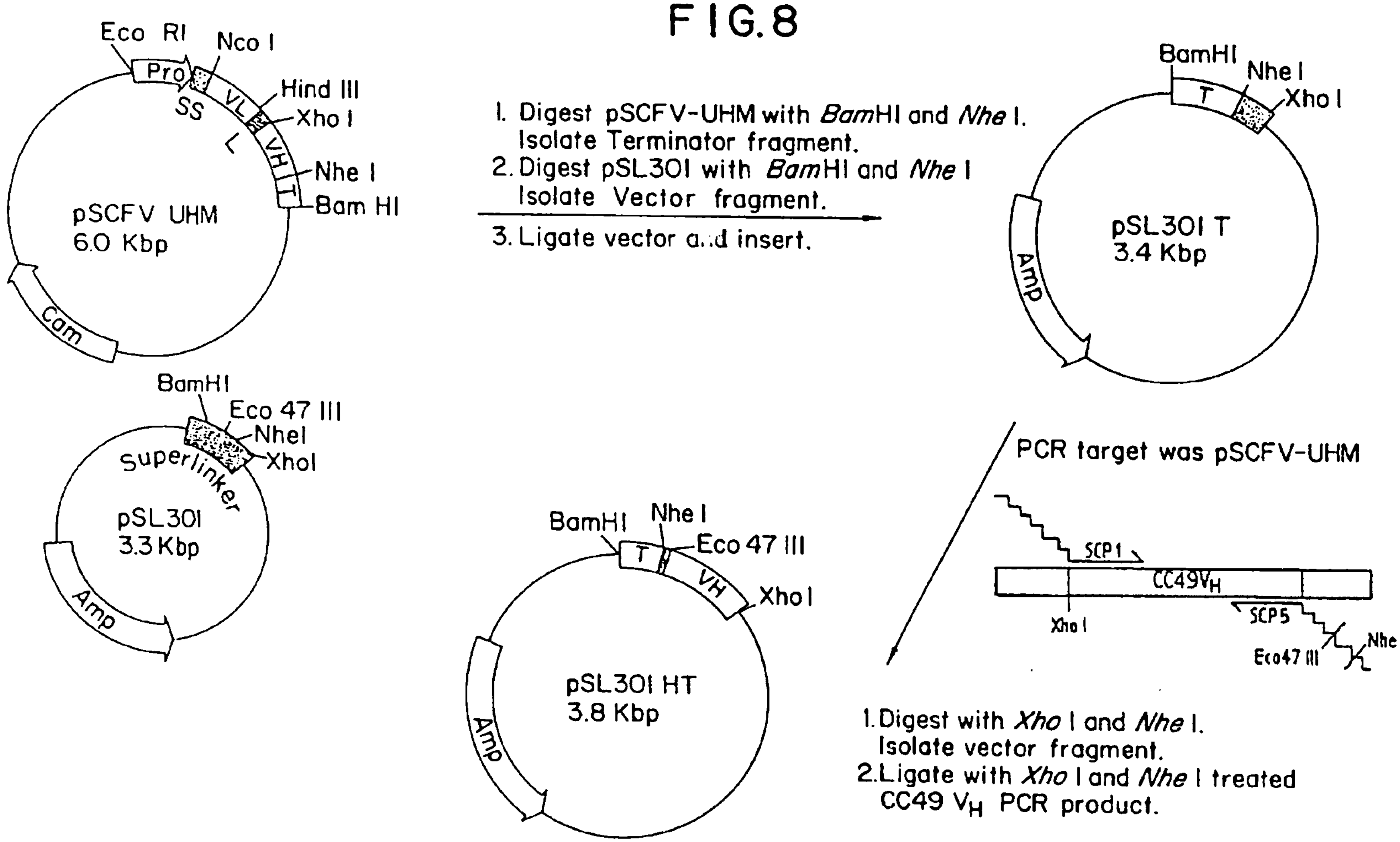
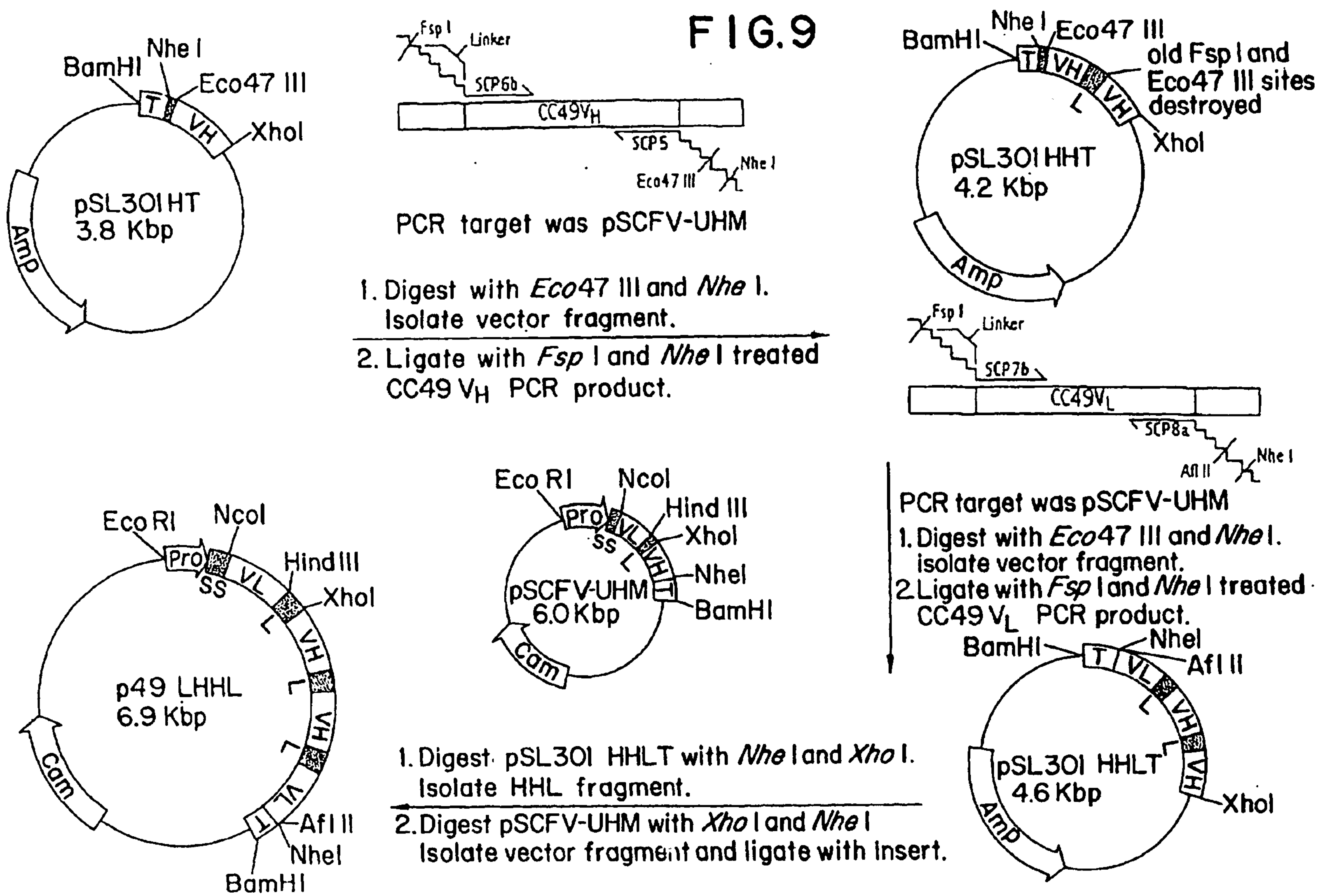

Eco RI
Nco I
Pro
SS
VL
L
Hind III
Xho I
VH
Nhe I
T
Bam HI
pSCFV UHM
6.0 Kbp
Cam
1. Digest pSCFV-UHM with BamHI and Nhe I.
Isolate Terminator fragment.
2. Digest pSL301 with BamHI and Nhe I
Isolate Vector fragment.
3. Ligate vector and insert.
BamHI
Nhe I
Xho I
T
pSL301 T
3.4 Kbp
Amp
BamHI
Eco 47 III
NheI
XhoI
Superlinker
pSL301
3.3 Kbp
Amp
PCR target was pSCFV-UHM
SCP1
CC49VH
Xho I
SCP5
Eco47 III
Nhe I
BamHI
Nhe I
Eco 47 III
T
VH
Xho I
pSL301 HT
3.8 Kbp
Amp
1. Digest with Xho I and Nhe I.
Isolate vector fragment.
2. Ligate with Xho I and Nhe I treated
CC49 VH PCR product.

FIG.9


BamHI
Nhe I
Eco47 III
XhoI
T
VH
pSL301HT
3.8 Kbp
Amp
Fsp I
Linker
SCP6b
CC49VH
SCP5
Eco47 III
Nhe I
PCR target was pSCFV-UHM
1. Digest with Eco47 III and Nhe I.
Isolate vector fragment.
2. Ligate with Fsp I and Nhe I treated
CC49 VH PCR product.
BamHI
Nhe I
Eco47 III
old Fsp I and
Eco47 III sites
destroyed
XhoI
pSL301HHT
4.2 Kbp
Amp
Fsp I
Linker
SCP7b
CC49VL
SCP8a
Afl II
Nhe I
PCR target was pSCFV-UHM
1. Digest with Eco47 III and Nhe I.
isolate vector fragment.
2. Ligate with Fsp I and Nhe I treated
CC49 VL PCR product.
BamHI
NheI
Afl II
XhoI
pSL301 HHLT
4.6 Kbp
Amp
Eco RI
NcoI
Hind III
XhoI
NheI
BamHI
Pro
SS
pSCFV-UHM
6.0 Kbp
Cam
1. Digest pSL301 HHLT with Nhe I and Xho I.
Isolate HHL fragment.
2. Digest pSCFV-UHM with Xho I and Nhe I
Isolate vector fragment and ligate with insert.
Eco RI
NcoI
HindIII
XhoI
Afl II
NheI
BamHI
p49 LHHL
6.9 Kbp
Cam

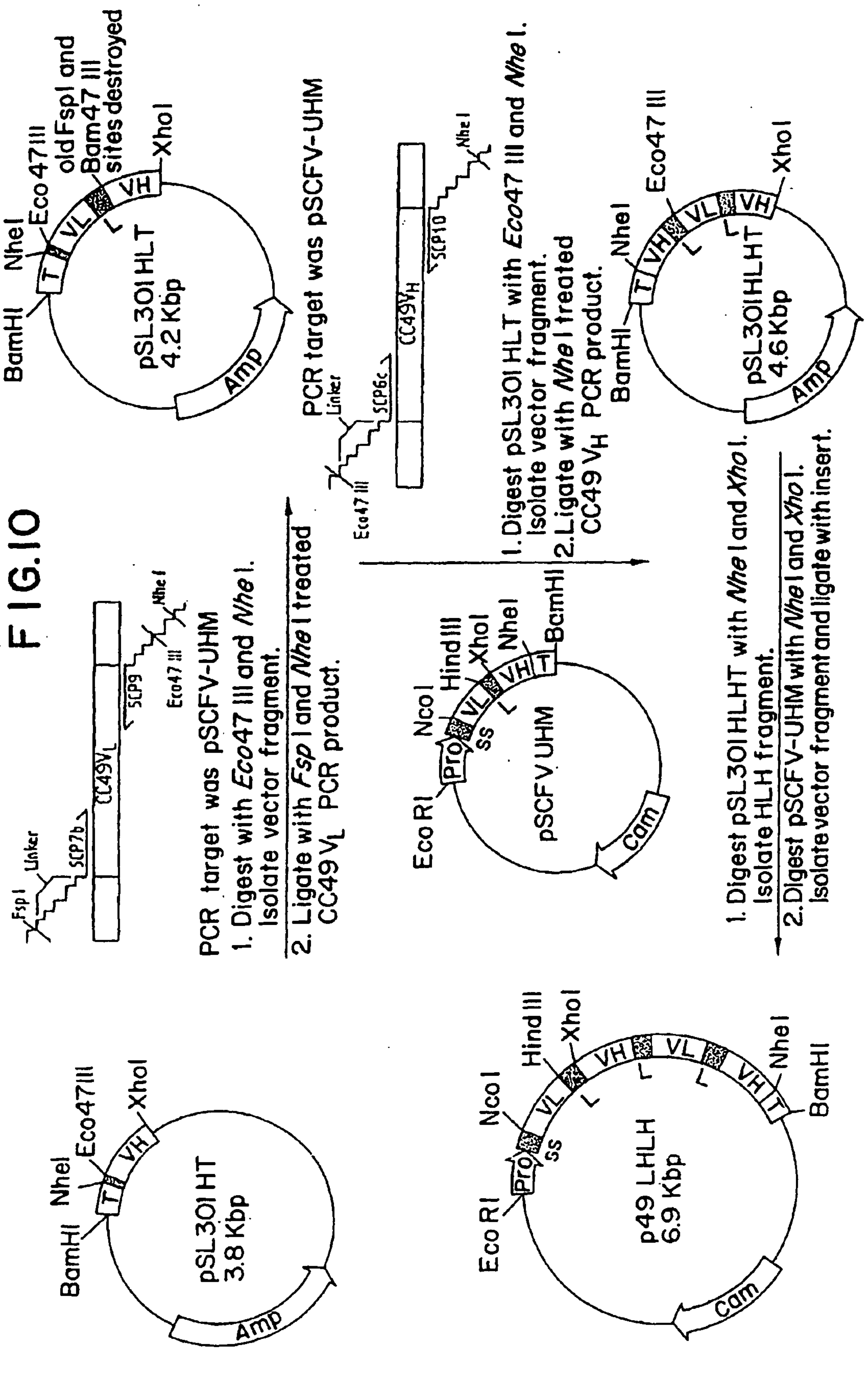
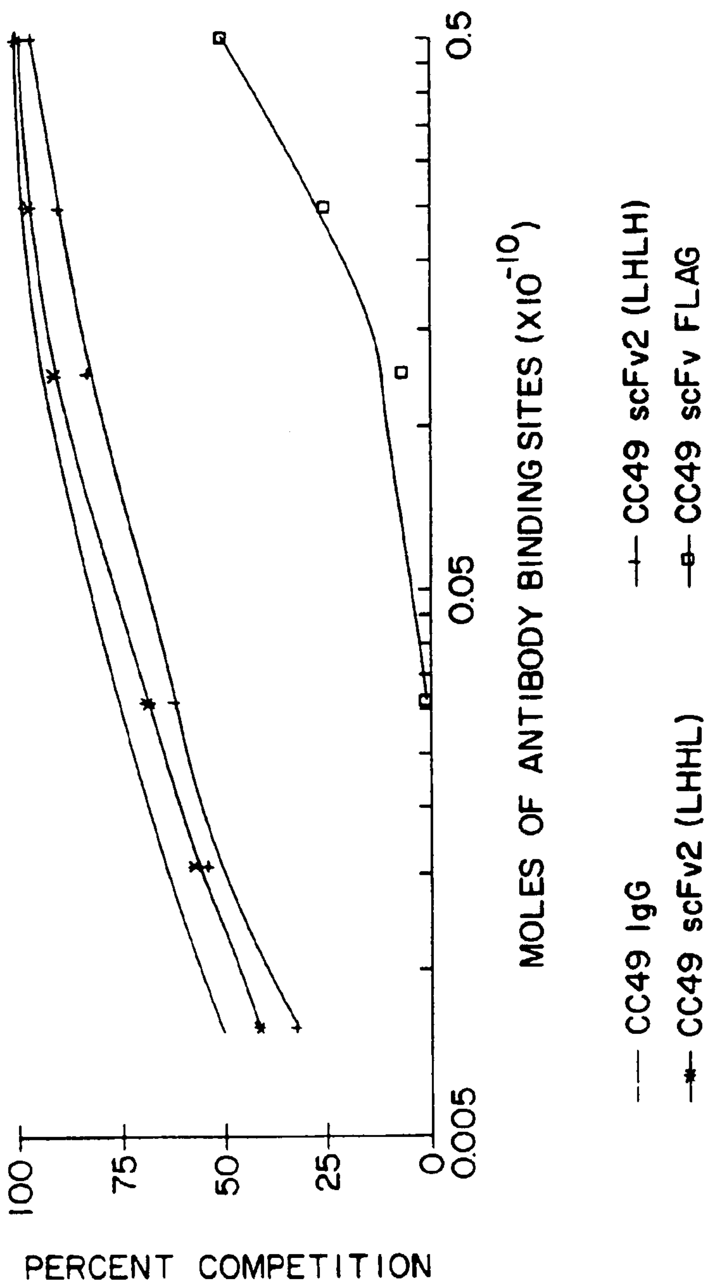

pSL301 HT
3.8 Kbp
BamHI
NheI
Eco47III
XhoI
T
VH
Amp
pSL301 HLT
4.2 Kbp
old FspI and Bam47 III sites destroyed
VL
L
Fsp I
Linker
SCP7b
CC49VL
SCP9
Eco47 III
Nhe I
PCR target was pSCFV-UHM
1. Digest with Eco47 III and Nhe I. Isolate vector fragment.
2. Ligate with Fsp I and Nhe I treated CC49 VL PCR product.
SCP6c
CC49VH
SCP10
1. Digest pSL301 HLT with Eco47 III and Nhe I. Isolate vector fragment.
2. Ligate with Nhe I treated CC49 VH PCR product.
pSL301 HLHT
4.6 Kbp
pSCFV UHM
EcoRI
NcoI
HindIII
Pro
SS
cam
1. Digest pSL301 HLHT with Nhe I and Xho I. Isolate HLH fragment.
2. Digest pSCFV-UHM with Nhe I and Xho I. Isolate vector fragment and ligate with insert.
p49 LHLH
6.9 Kbp

FIG.10

FIG.11


PERCENT COMPETITION
100
75
50
25
0
0.005
0.05
0.5
MOLES OF ANTIBODY BINDING SITES (X10^{-10})
CC49 IgG
CC49 scFv2 (LHHL)
CC49 scFv2 (LHLH)
CC49 scFv FLAG